

# Informationsgewinnung mittels Bindungsanalysen von Serumantikörpern an Peptidbibliotheken

Ein Ansatz zur serologischen Antikörper-Diagnostik, der ohne  
bekannte Antigene auskommt

## DISSERTATION

zur Erlangung des akademischen Grades  
doctor rerum naturalium  
(Dr. rer. nat.)  
im Fach Biologie

eingereicht an der  
Mathematisch-Naturwissenschaftlichen Fakultät I  
Humboldt-Universität zu Berlin

von

Frau Dipl. Natw. ETH Nicole Bruni  
geboren am 19.01.1973 in Zürich, Schweiz

Präsident der Humboldt-Universität zu Berlin:  
Prof. Dr. Christoph Marksches

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:  
Prof. Dr. Lutz-Helmut Schön

Gutachter:

1. Dr. Michal Or-Guil
2. Prof. Dr. Hanspeter Herzel
3. Prof. Dr. Sepp Hochreiter

eingereicht am: 04.08.2008  
Tag der mündlichen Prüfung: 19.12.2008



## Abstract

A characteristic trait of the humoral immune system is the production of lots of different antibodies. The repertoire of antibodies of any individual is of complex composition. Part of the antibody repertoire is generated without any influence from the environment, while another part is built solely due to external stimulations. Immune responses are the reactions of the body towards the antigenic stimulation from outside, or in case of autoimmunity from internal also. Antibodies with increasing antigen specificity over time are produced during immune responses. Such evolving antibodies oftentimes develop well-defined reactivities against the antigen that elicited the immune response. Commonly used diagnostic tests for the detection of serum-antibody-binding-properties successfully exploit these antibody reactivities against disease eliciting antigens. Such diagnostic tests do however need the knowledge of disease specific antigen as a prerequisite. Moreover, the tested antigen-antibody reactions should not cross-react with other antibody binders that are not connected to the tested disease. The search for diagnostic solutions in cases of missing known antigens and crossreactivities between healthy and diseased individuals might be solved by using the knowledge that serum antibody repertoires of diseased and healthy individuals have different recognition characteristics. The binding reactivities of whole antibody repertoires of different groups of individuals have different characteristic binding properties when tested against the antibody recognition of arbitrary probe molecule libraries. The arbitrary molecules of such libraries do not need to be related to the antigens of the disease to be diagnosed. The independence of the binding probe libraries to the disease-associated antigens is a promising approach for diagnosis of different diseases with the same arbitrary libraries. Peptides are the ideal molecules for such arbitrary binding probe libraries. The synthesis of peptides in large numbers is rather troublesome and expensive. For modern printing processes of glass chips arrays, though even small amounts of uniquely synthesized peptide may be used to produce hundreds of peptide carrier-chip replicas.

Based on the described premises the presented work demonstrates the differentiation of different diseases by means of antibody serum reactivity differences towards arbitrary peptide libraries between healthy and diseased individuals. The power of the diagnostic differentiation is shown by a combination of different technological implementations that include:

**peptide array technology** on cellulose membranes and glass chips,

**arbitrary peptide library** usage,

**serum incubation protocols** including antibody binding detection, visualisation and read-out techniques,

**normalisation** according to data distributions, and

**feature selection methods** for the extraction of the groups of the most differentiating peptides from high-dimensional peptide library spaces.

In further analysis, the resulting groups of feature-extracted peptides for diagnostic purposes are carefully analysed; the number of peptides necessary for reliable prediction are unanticipated small. Furthermore, the concept of »detecton« is introduced, discussed and conclusions are drawn there from. Surprisingly, »detectons« consist amongst others of peptides weakly recognized by all examined individuals. The low binding power as much as the small binding differences of the most differentiating peptides show that discriminating peptides do not need to be strong binders, nor do their differences in recognition need to be large in order to be relevant for diagnosis; a very important finding with potential impact on biomarker search. The unbiased approach of the library design works as well as it possibly could with intended libraries like whole-proteome arrays used in recent other works. Our findings shed a new light to the interpretation of the clinical relevance of differentially recognized peptides: a hasty interpretation deeming such peptides functional should be handled with care. We presume our approach to be an important step forward towards reliable diagnosis, in particular for diseases caused by yet unknown antigens. Furthermore, the high specificity of a »detecton« and its smallness in size (small number of members) might provide a basis for simultaneous diagnosis of various diseases on a single peptide microarray.

**Keywords:**

peptide array, antibody reactivity profiling, diagnosis, data analysis



## **Zusammenfassung**

Ein charakteristisches Merkmal des humoralen Immunsystems ist die Produktion vieler verschiedener Antikörper. Die Zusammensetzung des Repertoires der Antikörper ist äusserst komplex. Ein grosser Anteil des Antikörperrepertoires entsteht ohne Einfluss von äusseren Umweltfaktoren, während die Entstehung von anderen Teilen des Repertoires allein von der Stimulation durch äussere Reize abhängt. Immunantworten sind die Reaktionen des Körpers auf Antigen-Stimulationen von aussen (oder im Falle von Autoantigenen auch von innen). Antigen-Stimulationen führen zur Produktion von Antikörpern, deren Spezifität gegenüber dem auslösenden Antigen über die Zeit der Immunantwort zunimmt. Geläufige diagnostische Tests für den Nachweis von Krankheit-spezifischen Serumantikörpern nutzen diese Antigene zum Nachweis der Krankheit via Antikörperbindung. Derartige diagnostische Tests setzen jedoch die Kenntnis von Krankheit-spezifischen Antigenen voraus. Darüber hinaus dürfen die gesuchten Antigen-Antikörper-Bindungen keine Kreuzreaktivität zu anderen, nicht Krankheits-assoziierten Antikörpern aufweisen. Die Bindungsreaktivitäten ganzer Antikörperrepertoires verschiedener Gruppen von Individuen haben unterschiedliche Eigenschaften. Dies wird anhand der Messung der Antikörper-Serensamples gegenüber beliebiger Molekül-Bibliotheken gezeigt. Die Moleküle dieser beliebigen (Zufalls-) Bibliotheken müssen dabei keine Verwandtschaft zu den Antigenen der Krankheit haben. Die Unabhängigkeit der Bindungspartnerbibliotheken von den assoziierten Antigenen ist ein vielversprechender Ansatz zur Diagnose verschiedener Krankheiten mittels derselben zufälligen Bibliotheken. Ideale Moleküle zur Verwendung in Zufallsbibliotheken sind Peptide, deren Synthese in grosser Zahl aufwendig und teuer ist. Mit modernen Herstellungsverfahren von Peptidarrays auf Glasträgern können mit kleinen, einmal synthetisierten Peptidmengen hunderte von Träger-Replikas produziert werden. Die Suche nach hochaffinen Bindern zur Diagnose von Krankheiten mit unbekannten Antigenen oder mit Kreuzreaktivitäten zwischen gesunden und kranken Individuen könnte überflüssig werden durch die Erkenntnis, dass die Serumantikörper-Repertoires von kranken und gesunden Individuen unterschiedliche Bindungseigenschaften besitzen.

Gestützt auf die beschriebenen Voraussetzungen zeigt die vorliegende Arbeit, dass die Messung von Serumantikörper-Bindungen gegenüber Peptidbibliotheken mit zufälligen Sequenzen die Unterscheidung zwischen Gruppen

von gesunden und kranken Individuen für unterschiedliche Krankheiten ermöglicht. Eine Kombination verschiedener Technologien ermöglicht die hohe Zuverlässigkeit der diagnostischen Differenzierung. Die wichtigsten Technologien sind:

**Peptidarrays** auf Zellulosemembranen und auf Glasträgern,

**zufällige Peptidbibliotheken** als Proben,

**Serumantikörper-Inkubations-Protokolle** inklusive Detektion und Visualisierung (bzw. Ausleseverfahren) der Antikörperbindungen,

**Normierungen** entsprechend der Datenverteilung,

**Merkmalsselektion** zur Bestimmung von Peptidgruppen aus einem hochdimensionalen Peptidraum, die zwischen krank und gesund unterscheiden können.

Des Weiteren werden die resultierenden Gruppen, der mittels Merkmalsselektion gewählten Peptide, für diagnostische Zwecke sorgfältig analysiert; eine unerwartet kleine Anzahl von Peptiden ist ausreichend für eine zuverlässige Vorhersage der untersuchten Gruppen. Ausserdem wird das Konzept des »Detekton« eingeführt und diskutiert. Ein nicht vorhersehbares Resultat war des Weiteren, dass die Detektone Peptide enthalten, die von allen untersuchten Seren nur schwach erkannt werden. Die Klassifizierung unter Beteiligung kleiner Bindungsstärken und ebensolcher Bindungsstärkenunterschieden zeigt, dass diagnostisch relevante Peptide keine hohen Binder sein müssen – für die Suche nach Biomarkern eine sehr wichtige Erkenntnis. Der unvoreingenommene Ansatz mit Zufallsbibliotheken ermöglicht eine ebenso gute Unterscheidung von gesund und krank, wie sie auch mit voreingenommenen Bibliotheken (z. B. Arrays mit vollständigen Proteomen) gezeigt worden sind. Unsere Ergebnisse werfen ein neues Licht auf die Interpretation der klinischen Relevanz von differenziell erkannten Peptiden. Eine voreilige Interpretation von stark bindenden Peptiden sollte sorgfältig geprüft werden. Wir vermuten, dass unser Ansatz ein wichtiger Schritt in Richtung zuverlässiger Diagnose darstellt, insbesondere für Krankheiten mit noch unbekannten Antigenen. Ausserdem bietet die hohe Spezifität der Detektone und deren kleinen Mitgliederzahl eine Grundlage für die gleichzeitige Diagnose von verschiedenen Krankheiten auf einem einzigen Peptid-Mikroarray.

#### **Schlagwörter:**

Peptidarray, Antikörper-Reaktivitäts-Profile, Diagnose, Datenanalyse

FÜR BEN UND LEONARD



# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>2</b>
1.1	Immunsystem und Immunreaktionen . . . . .	2
1.1.1	Das angeborene Immunsystem . . . . .	2
1.1.2	Das adaptive Immunsystem . . . . .	3
1.2	Die humorale Immunantwort . . . . .	3
1.2.1	Differenzierung von B-Zellen und Entstehung von Antikörpern . . . . .	4
1.2.2	Struktur von Antikörpern . . . . .	5
1.2.3	B-1-Zellen und »natürliche Antikörper« . . . . .	6
1.2.4	Das Antikörperrepertoire . . . . .	7
1.2.5	Konzentrationen von Lymphozyten und Antikörpern . . . . .	8
1.2.6	Spezifität, Diversität und Kreuzreaktivität von Antikörpern . . . . .	8
1.2.7	Das Protektion . . . . .	9
1.3	Immunantworten gegenüber Pathogenen . . . . .	10
1.3.1	Nematoden und Nematodeninfektionen . . . . .	10
1.3.2	Pertussis . . . . .	10
1.4	Die Charakterisierung von monoklonalen Antikörpern und Serumantikörpern . . . . .	11
1.4.1	Charakterisierung monoklonaler Antikörper . . . . .	11
1.4.2	Charakterisierung von Serumantikörper . . . . .	12
1.5	Technologien der Bindungsdetektion mittels Hochdurchsatzstrategien . . . . .	14
1.5.1	Probemoleküle: Peptide eignen sich besser als Proteine . . . . .	14
1.5.2	Peptidbibliotheken . . . . .	15
1.5.3	Peptidsynthese: Die SPOT Synthese . . . . .	16
1.5.4	Trägermaterialien: Makro- und Mikroarrays . . . . .	17
1.6	Serumantikörper-Reaktivitäts-Profile auf Mikroarrays in der Literatur . . . . .	18
1.7	Dimensionsreduktion . . . . .	19

1.7.1	Unvoreingenommene Dimensionsextraktion . . . . .	19
1.7.2	Merkmalsselektion . . . . .	20
1.8	Mathematische Modellierung von Wechselwirkungen zwischen Antikörpern und Antigenen . . . . .	20
1.8.1	Das Konzept des »Shape Space« . . . . .	20
<b>2</b>	<b>Zielsetzung</b>	<b>24</b>
<b>3</b>	<b>Material und Methoden</b>	<b>26</b>
3.1	Mäuse und Serengewinnung . . . . .	26
3.1.1	Mäuse . . . . .	26
3.1.2	<i>H.-polygyrus</i> -Infektionen . . . . .	26
3.1.3	Serengewinnung . . . . .	27
3.2	Peptidbibliotheken . . . . .	27
3.2.1	Design von Peptidbibliotheken mit zufälligen Aminosäuresequenzen . . . . .	27
3.3	Peptidsynthese und Herstellung der Peptidarrays . . . . .	28
3.3.1	SPOT Synthese . . . . .	28
3.3.2	Herstellung von Peptidarrays auf Zellulosemembranen (Makroarrays) . . . . .	29
3.3.3	Herstellung von Peptidarrays auf Glasträgern (Mikroarrays) . . . . .	30
3.4	Serumantikörper-Bindungsmessung auf Zellulosemembranen (Makroarrays) . . . . .	31
3.5	Serumantikörper-Bindungsmessung auf Glasträger (Mikroarrays) . . . . .	31
3.6	Datenakquise und Auslesung . . . . .	32
3.6.1	Signal-Rausch-Verhältnis . . . . .	32
3.7	Normierung der Daten . . . . .	33
3.8	Extern akquirierte Daten . . . . .	34
3.8.1	Pertussis-Seren . . . . .	34
3.9	Datenanalyse und Statistik . . . . .	34
3.9.1	Hauptkomponentenanalyse (PCA) . . . . .	34
3.9.2	Lineare Diskriminanten Analyse (LDA) . . . . .	34
3.9.3	Anpassung an die Normalverteilung . . . . .	34
3.9.4	Korrelation . . . . .	35
3.9.5	Sensitivität und Spezifizität . . . . .	35
3.9.6	Wilcoxon-Vorzeichen-Rang-Tests . . . . .	35
3.9.7	Heat maps und Clustering . . . . .	35
3.9.8	Multidimensionale Skalierung . . . . .	36
3.9.9	Potential Support Vector Machine (P-SVM) . . . . .	36

3.9.10	Bestimmung der Signifikanzen der Resultate . . . . .	37
<b>4</b>	<b>Optimierung und Evaluation der Technologien</b>	<b>38</b>
4.1	Übersicht der verwendeten Technologien und Prozesse . . . . .	39
4.2	Grundlagen der Technologien . . . . .	41
4.2.1	Detektion von Serumantikörperbindungen . . . . .	41
4.2.2	Genauigkeit der Bindungsmessungen: Binder und Nicht- Binder . . . . .	43
4.2.3	Synthese von Cystein-enthaltenden Peptiden . . . . .	44
4.2.4	Regeneration der Zellulosemembranen . . . . .	44
4.2.5	CAPE-Membranen und Beta-Ala-Membranen im Ver- gleich . . . . .	45
4.2.6	Grössere Zahl identischer Träger pro Charge – Reduk- tion der Anzahl der Peptide . . . . .	45
4.2.7	Verzerrung der experimentellen Ergebnisse durch Ver- suchsanordnung . . . . .	47
4.3	Reproduzierbarkeit der Serumantikörperbindungs-Signale an Peptidarrays auf Zellulosemembranen . . . . .	47
4.4	Reproduzierbarkeit der Serumantikörperbindungs-Signale an Peptidarrays auf Glasträgern . . . . .	49
4.4.1	Unbeständigkeit der Signale von Glasträgern . . . . .	50
4.4.2	Die Reproduzierbarkeit qualitativ-hochwertiger Glas- träger . . . . .	52
4.4.3	Signalintensitäten-Verteilungen der Glasträgerdaten . . . . .	55
4.5	Auslesen der Signalintensitäten . . . . .	55
4.5.1	Software zum Auslesen der Signalintensitäten . . . . .	55
4.5.2	Signalintensitäten von Peptidspots und Hintergrund im Vergleich . . . . .	57
4.5.3	Korrektur der sekundären Antikörper-Signale . . . . .	59
4.6	Peptid-Peptid Bindungsversuch auf Glasträgern . . . . .	60
<b>5</b>	<b>Resultate Zellulosemembran</b>	<b>62</b>
5.1	Hauptkomponentenanalyse und Normierung der Daten . . . . .	62
5.2	Grenzen der Hauptkomponentenanalyse . . . . .	64
5.3	LDA, Dimensionsreduktion und Merkmalsselektion . . . . .	65
5.4	LDA-Multiklassen-Klassifizierung . . . . .	67
5.5	Merkmalsselektion versus Normierung . . . . .	68
<b>6</b>	<b>Resultate Glasträger</b>	<b>70</b>
6.1	Verzicht auf Normierung der Daten . . . . .	70
6.2	Hauptkomponentenanalyse . . . . .	72

6.3	eduktion der Zahl analysierter Gruppen . . . . .	73
6.4	Paarweise Klassifizierung von drei Gruppen mittels Merkmals- selektion . . . . .	75
6.5	Vom Protekton zum Detekton . . . . .	76
6.5.1	Definition Detekton . . . . .	76
6.5.2	Detektone sind überall . . . . .	77
6.6	Merkmalsselektion: P-SVM versus Wilcoxon-Vorzeichen-Rang- Test . . . . .	79
6.7	Multiklassen-Klassifizierung in zwei Dimensionen . . . . .	84
<b>7</b>	<b>Klassifizierung humaner Serumantikörper</b>	<b>88</b>
7.1	Klassifizierung von humanen Seren von Pertussis-geimpften und Placebo-behandelten Individuen . . . . .	88
<b>8</b>	<b>Schlussfolgerungen und Ausblicke</b>	<b>92</b>
8.1	Erreichte Ziele . . . . .	92
8.2	Zusammensetzung der Probenbibliotheken . . . . .	93
8.3	Normierung und Nicht-Normierung . . . . .	94
8.4	Diagnose und Infektionsverlauf . . . . .	95
8.5	Vergleich der Inkubationsträger Zellulosemembranen und Glas- träger – Makro versus Mikro . . . . .	96
8.5.1	Vergleich verschiedener Methoden der Herstellung von Zellulosemembranen . . . . .	97
8.6	Extern akquirierte Datensätze . . . . .	97
8.7	Protektone und Detektone . . . . .	98
8.7.1	Weitere Untersuchungen der Detektone . . . . .	99
8.7.2	IgM- und IgG-Detektone im Vergleich . . . . .	99
8.7.3	IgM- und IgG-Antikörperbindungsmessungen im Ver- gleich . . . . .	100
8.8	Shape Space und Modellierung . . . . .	101
8.9	Mathematische Modellierung der Antikörpererkennung . . . .	103
8.9.1	Das Bitstring-Modell von Lancet . . . . .	103



# Abbildungsverzeichnis

4.1	Übersicht: Serumantikörper-Inkubationen und Analysen der resultierenden Bindungssignale auf Peptidbibliotheken . . . . .	40
4.2	Sortierte Signalintensitäten eines Serenpools auf einem ersten Membranstück inkubiert . . . . .	43
4.3	Mittels LDA-Richtung selektierte Peptide . . . . .	46
4.4	Mittelwerte und Standardabweichungen der Signalintensitäten von 156 Peptiden von jeweils drei Inkubations-Repetitionen . .	49
4.5	Die Bilder der inkubierten Zellulosemembranen aus der zweiten Charge . . . . .	51
4.6	Die Verteilung der IgM-Signale auf einer Zellulosemembran . .	52
4.7	Peptidarrays auf Glasträgern mit vielen Repetitionen und von unterschiedlicher Qualität . . . . .	53
4.8	Die Qualitätszunahme der Inkubation von Seren auf Glasträgern anhand von zwei Beispiel-Subarrays . . . . .	54
4.9	Mittelwerte $\mu$ und Standardabweichungen $\sigma$ der SI von 15 Inkubations-Repetitionen auf Glasträgern . . . . .	54
4.10	Signalintensitäten-Verteilung auf Glasträgern . . . . .	56
4.11	Logarithmierte Signalintensitäten von 15 Repetitionen und deren Gaussverteilung . . . . .	56
4.12	Signalintensitäten-Verteilung aller BALB/c-Inkubationen auf Glasträgern . . . . .	57
4.13	Inkubierte Zellulosemembranen aus Charge fünf . . . . .	58
4.14	Abhängigkeit der Signalintensitäten der Peptidspots von den Hintergrundsignalen . . . . .	59
5.1	Die Mäusegruppen, deren Seren auf Zellulosemembranen inkubiert und analysiert wurden . . . . .	63
5.2	Hauptkomponentenanalysen und Normierung der Inkubationen von Seren auf Zellulosemembranen . . . . .	64
5.3	Hauptkomponentenanalyse der IgM-Antikörpersignale von auf Zellulosemembranen inkubierten Seren . . . . .	65

5.4	LDA Klassifizierung des Zellulosemembrandatensatzes – mit und ohne Merkmalsselektion . . . . .	67
5.5	Zweidimensionale LDA-Klassifizierung der SI von BALB/c-Seren auf Zellulosemembranen inkubiert . . . . .	69
6.1	Mäusegruppen, deren Seren auf Glasträger inkubiert und analysiert wurden . . . . .	71
6.2	Normierung und Additionseffekte der P-SVM-selektierten Peptide . . . . .	73
6.3	Hauptkomponentenanalyse (PCA) der IgM- und IgG-Antikörpersignale von auf Glasträgern inkubierten Seren . . . . .	74
6.4	Die ersten drei Detektone zur Vorhersage der »gesunden« gegenüber der »früh nach Infektion« Gruppen . . . . .	78
6.5	Übersicht der Eigenschaften des IgM-Ak-Datensatzes auf Glasträgern . . . . .	80
6.6	Übersicht der Eigenschaften des IgG-Ak-Datensatzes auf Glasträgern . . . . .	81
6.7	Eigenschaften der Detektone der IgM-Antikörperbindungen . .	82
6.8	Eigenschaften der Detektone der IgG-Antikörperbindungen . .	83
6.9	Vergleich der Merkmalsselektion-Methoden Wilcoxon-Test und P-SVM . . . . .	85
6.10	Zweidimensionale LDA-Multiklassen-Klassifizierung . . . . .	86
7.1	Hauptkomponentenanalyse der Signalintensitäten von Pertussis-geimpften und Placebo-behandelten Seren . . . . .	90
7.2	Clusteranalyse der Signalintensitäten von Pertussis-geimpften und Placebo-behandelten Seren . . . . .	91
8.1	Simulierte Affinitätsverteilungen von neun verschiedenen Bitstring-Antikörpern mit verschiedener Anzahl von 1en . . . . .	105
8.2	Die simulierten Affinitäten einer Schritt-für-Schritt Antikörper-Mutation im dreidimensionalen Raum . . . . .	106

# Tabellenverzeichnis

1.1	Übersicht der Eigenschaften von Makro- und Mikroarrays . . .	18
3.1	Verteilung der Aminosäuren auf der Oberfläche von Proteinen	28
4.1	Übersicht der Zellulosemembran-Chargen . . . . .	48
4.2	Korrelationskoeffizienten der Membranen der zweiten und vier- ten Charge . . . . .	50
5.1	Klassifizierung und Kreuzvalidierungs-Vorhersage des Zellulo- semembrandatensatzes . . . . .	68
6.1	Sensitivität und Spezifität der Klassifizierung der gesunden und <i>H.-polygyrus</i> -infizierten BALB/c-Mäusen der Glasträger- datensätze . . . . .	75
6.2	Die Sequenzen der Detekton-Merkmale zwischen allen klassi- fizierten Mäusegruppen-Paaren auf Glasträgern . . . . .	76
6.3	Peptide, die von P-SVM selektiert wurden für die höchsten diagnostischen Zuverlässigkeiten zwischen gesunden BALB/c- Mäusen und <i>H.-polygyrus</i> -infizierten spät nach Infektion . . .	84
8.1	Die Bitstring-Antikörper . . . . .	107



# Kapitel 1

## Einführung

### 1.1 Immunsystem und Immunreaktionen

Lebewesen mit einem Immunsystem sind in der Lage, potenzielle Fremdkörper (d.h. Antigene) zu erkennen und sie mittels einer Immunantwort zu bekämpfen. Die Hauptaufgabe des Immunsystems ist die Verteidigung des Organismus gegen infektiöse Eindringlinge. Eindringlinge sind Viren, einzellige Bakterien und Pilze oder auch mehrzellige Parasiten wie etwa Bandwürmer. Eine weitere erwähnenswerte Aufgabe ist die Erkennung und Eliminierung von krankhaft entarteten körpereigenen Zellen, die z. B. zu Krebs führen können. Das Immunsystem setzt sich aus physiologischen Barrieren, löslichen Stoffen und Zellen zusammen, die im ganzen Körper verteilt sind, um Antigene davon abzuhalten einzutreten, sich zu vermehren und Schaden anzurichten. Als Antigen bezeichnet man alles was eine Immunreaktion auslösen kann. Das Immunsystem besteht aus angeborenen (engl. *innate*) und adaptiven Abwehrmechanismen.

#### 1.1.1 Das angeborene Immunsystem

Bereits evolutionär primitive Lebewesen weisen Komponenten des angeborenen Immunsystems auf. Die Mechanismen der angeborenen Immunität sind, im Gegensatz zu denen der adaptiven Immunität, nicht in der Lage, auf wiederholt auftretende Antigene im Verlaufe der Zeit schneller zu reagieren; sie können kein immunologisches Gedächtnis aufbauen. Die Gene der angeborenen Immunität sind unveränderlich codiert, und die Genprodukte sind auf spezielle Rezeptoren oder Signale der Antigene ausgerichtet. Die angeborene Immunität umfasst Epithel-Barrieren (mechanische, chemische und mikrobiologische), Fress- und Killerzellen (Makrophagen, Granulozyten, natürliche Killerzellen und neutrophile Zellen), entzündungsauslösende Mecha-

nismen und das Komplementsystem. Des Weiteren stimuliert das angeborene Immunsystem das adaptive Immunsystem. Eine detailliertere Beschreibung des angeborenen Immunsystems findet sich in Lehrbüchern [32], [1]. Je nach Definition gehören zur angeborenen Immunität auch die B-1-Zellen (bzw. B-1-Lymphozyten). Diese sind eine Untergruppe der B-Zellen, die per Definition [32] jedoch zur adaptiven Immunität gehören. Eine genauere Beschreibung der B-1-Zellen und deren Funktionen folgt in Abschnitt 1.2.3; dort werden sie bezüglich ihrer Bedeutung in dieser Arbeit näher behandelt.

### 1.1.2 Das adaptive Immunsystem

Die adaptive bzw. erworbene Immunabwehr ist nur bei Wirbeltieren vorhanden und besteht aus den T- und B-Zellen (d.h. den Lymphozyten) sowie den Antigenpräsentierenden dendritischen Zellen. Die B- und T-Zellen durchlaufen erste Reifungsprozesse im Thymus und im Knochenmark, in deren Verlauf bereits eine grosse Diversität von Lymphozyten entsteht. Nur ein Teil der Zellen überlebt die Qualitätskontrollen nach den ersten Reifungsprozessen. Die Überlebenden werden in den Organismus abgegeben. Somatische Diversität wird später in weiteren Reifungsprozessen in spezialisierten Geweben wie den Lymphknoten und der Milz zugefügt, wodurch die Lymphozyten, die anfänglich schlechter erkannten Antigene mit steigender Affinität zu binden vermögen. Generell zeichnet sich die adaptive Immunität durch hohe Diversität und Plastizität aus. Bei der Antigenerkennung des adaptiven Immunsystems unterscheidet man drei Klassen von Rezeptoren:

**Haupthistokompatibilitätskomplexe** bzw. *MHC*-Rezeptoren (von engl. *Major Histocompatibility Complex*); es gibt zwei Klassen von *MHCs*: *MHC I* und *MHC II*, wovon Erstere von allen körpereigenen Zellen exprimiert werden.

**T-Zell-Rezeptoren** bzw. *TCR* (von engl. *T cell receptor*), die von T-Zellen exprimiert werden und Antigene erkennen, die von *MHCs* präsentiert werden.

**B-Zell-Rezeptoren** bzw. *BCR* (von engl. *B cell receptor*), die von B-Zellen exprimiert werden und entweder auf deren Oberflächen präsentiert werden oder von ihnen als lösliche Antikörper sezerniert werden.

## 1.2 Die humorale Immunantwort

Der Begriff der »humoralen« Immunantwort stammt vom lateinischen Wort *humor*, das Flüssigkeit bedeutet. Die Theorie der *humores* lässt sich zu alten

griechischen Philosophen zurückverfolgen. Die Humorallehre betrachtet die Gesundheit als Balance von vier Körperflüssigkeiten (bzw. *humores*).

### 1.2.1 Differenzierung von B-Zellen und Entstehung von Antikörpern

Antikörper (Ak) sind Proteine und haben eine Y-förmige Struktur von etwa 150 kDa Molekulargewicht. Antikörper werden auch Immunglobuline (Ig) genannt und sind die löslichen, sezernierten Produkte von B-Zellen. B-Zellen werden auch B-Lymphozyten genannt. B-Zellen können zu Antikörper-sezierenden Plasma- oder Gedächtnis-Zellen reifen, nachdem sie durch Erkennen eines Antigens aktiviert wurden. Die Produktion und Sekretion von Antikörpern ist das Hauptmerkmal des humoralen Immunsystems. Jede B-Zelle produziert nur einen *BCR*- oder Antikörper-Klon von einer Spezifität, eine Tatsache die bereits 1958 von Nossal [49] entdeckt wurde. Die Antikörper, die von einer B-Zelle produziert werden, haben identische Aminosäuresequenzen, die von komplizierten, genau regulierten und gut kontrollierten Reifungsprozessen der B-Zelle und deren DNA abhängen. Der Verlauf der Reifungsprozesse aller B-Zellen zusammen und die davon abhängigen Mengen und Diversitäten des effektiven Antikörperrepertoires (siehe Abschnitt 1.2.4) teilen sich in einen Antigen-unabhängigen und einen Antigen-abhängigen Teil auf. Dabei sind nur Plasmazellen in der Lage, Antikörper zu sekretieren. Den Bindungseigenschaften der Antikörper gegenüber Antigenen kommt eine wichtige Rolle beim Erkennen, Neutralisieren und Zerstören von eindringenden Pathogenen zu. Andererseits spielen Antikörper in Form von Autoantikörpern eine entscheidende Rolle bei Autoimmunerkrankungen. Und letztlich sind Antikörper auch in Allergieerkrankungen involviert, welche als weitere unerwünschte Überreaktionen des Immunsystems auftreten können.

Zur Verteidigung des Organismus ist es entscheidend, dass das effektive Antikörperrepertoire möglichst divers ist. Die Diversität ist erforderlich, weil jeder potenziell gefährliche Eindringling erkannt werden muss. Der Antigen-unabhängige Reifungsprozess findet im Knochenmark statt. Die daraus resultierende Diversität hängt zunächst stark vom genetischen Hintergrund des Individuums ab. Bei der Reifung im Knochenmark fügen auch Umlagerungen und Rekombinationen der Antikörper-codierenden Gene stochastische Variabilität bei. Und schliesslich führen Selektions-Prüfstellen im Knochenmark zu einer Auswahl von B-Zellen, die überlebenstaugliche Antikörper codieren [84]. Nur B-Zellen, die diese ersten drei Stufen der Selektion überstanden haben, sind potenziell in der Lage, am adaptiven Teil des Immunsystems

teilzunehmen und dürfen das Knochenmark überhaupt verlassen. Ab diesem Punkt bezeichnet man sie als reife B-Zellen. Reife B-Zellen, die ein Antigen zu binden vermögen, können dann unter Umständen einen weiteren Reifungsprozess in sogenannten Keimzentren durchlaufen. Die Keimzentrumsreaktion ist abhängig von T-Zellen und deren Antigen-Präsentation. Keimzentrumsreaktionen finden innerhalb der Keimzentren von lymphoiden Organen statt, wo die B-Zellen erneute Mutationen ihrer Antikörpercodierenden DNA durchlaufen und aufgrund ihrer gesteigerten oder abfallenden Bindungsfähigkeit gegenüber dem präsentierten Antigen positiv oder negativ selektiert werden. Positiv selektierte B-Zellen teilen sich schnell und mutieren bei jeder Teilung erneut. Dabei ändern sich (d.h. mutieren) nicht nur die Antigen-Bindungsstellen (d.h. CDR-Regionen, siehe Abschnitt 1.2.2) der B-Zellrezeptoren. Auch das Fc-Fragment, welches den Isotypen des B-Zellrezeptors oder Antikörpers bestimmt, kann als Ganzes ersetzt werden. Jeder Antikörper gehört aufgrund seines Fc-Fragmentes zu einer der fünf Klassen (d.h. Isotypen), die IgM, IgG, IgA, IgE und IgD heißen. Ein Austausch des Fc-Rezeptors kommt daher einem Klassenwechsel gleich. Der Isotyp bestimmt die Reaktionen und Effekte, die ausgelöst werden, nachdem der betreffende Antikörper ein Antigen erkannt hat. Eine gute Übersicht über die Fc-Rezeptor-abhängigen Effekte der Fc $\gamma$ -Rezeptoren liefern Nimmerjahn u. Ravetch [47].

B-Zellen und deren Antikörper, die aus der Keimzentrumsreaktion hervorgehen, binden die Antigene, gegenüber denen sie gereift sind, mit verbesserter Affinität. Aus diesen B-Zellen resultierende Plasma- und Gedächtniszellen sind Teil des immunologischen Gedächtnisses, dank derer ein Individuum besser und schneller auf wiederholt auftretende Antigene reagieren kann. Plasmazellen werden zudem in lang- und kurzlebige Plasmazellen [58] unterteilt, was spätestens bei der Interpretation von Bindungsmessungen von Antikörperrepertoires, die zu unterschiedlichen Zeiten gewonnen wurden, bedacht werden sollte. Plasmazellen sind in der Lage, bis zu 1000 Antikörper pro Sekunde zu synthetisieren und zu sezernieren. Eine Menge, die auch angesichts von IgM und IgG Konzentrationen von 1 – 10 Gramm Antikörper pro Liter in humanem Plasma [73] zu schnellen Veränderungen der Zusammensetzung des Antikörperrepertoires führen kann.

## 1.2.2 Struktur von Antikörpern

Die beiden Äste der Y-Struktur von Antikörpern können mit dem Enzym Papain abgespalten werden. Die abgespaltenen Äste der Y-Struktur heißen Fab-Fragmente. Die Fab-Fragmente enthalten die hypervariablen CDR-



Regionen, welche die Antigene binden. Die CDR1-, CDR2- und CDR3-Regionen der Antikörper-DNA werden in den Keimzentrumsreaktionen am stärksten verändert und mutiert. Der Stamm der Y-Struktur wiederum heisst Fc-Fragment und bestimmt den Isotyp und damit die Effektorfunktion des Antikörpers. Fc-Fragmente werden nur vollständig ausgetauscht, was einem Klassenwechsel und damit einer Änderung der Effektorfunktion gleichkommt. Eine ausführliche Beschreibung der Struktur und der Effektorfunktionen von Antikörpern findet sich in Janeway [32] und Nimmerjahn u. Ravetch [47].

### 1.2.3 B-1-Zellen und »natürliche Antikörper«

In verschiedenen Experimenten wurde gezeigt, dass auch Tiere, die ihr ganzes Leben ohne Kontakt zu antigenen Substanzen verbracht haben, erhebliche Mengen von Antikörpern produzieren [53]. Antikörper, die ohne Antigenkontakt bzw. ohne vorhergehende Immunantwort entstehen, nennt man »natürliche Antikörper«. Die Definitionen der natürlichen Antikörper sind unterschiedlich. Die meisten Publikationen referenzieren die Definition der natürlichen Antikörper auf Jerne [33], Boyden [11] oder Avrameas [5]. In den Abschnitten 5 und 6 dieser Arbeit werden verschiedene IgM-Repertoires analysiert, zu denen natürliche Antikörper vermutlich entscheidend beitragen. Folglich werden hier vorab die wichtigsten bekannten Eigenschaften der natürlichen Antikörper zusammengetragen:

Natürliche Antikörper der IgM-Klasse stammen von B-1-Zellen ab. Die Existenz der B-1 Zellen ist allerdings umstritten. Definiert werden B-1-Zellen als eine Untergruppe der B-Zellen, die gekennzeichnet ist durch die Expression der CD5-Zelloberflächen-Proteine (wobei es je nach Definition auch B-1-Zellen ohne CD5 gibt, bzw. die B-1-Zellen in Ratten generell CD5-frei sind [32]). Die B-1-Zellen entstehen früh in der Ontogenese und produzieren nur ein eingeschränktes Repertoire von Antikörpern, was auf die beschränkte Kombination von Genumlagerungen und eine Absenz von Hypermutationen zurückzuführen ist. B-1-Zellen können ohne T-Zellhilfe eine Immunantwort starten und IgM Antikörper sezernieren. Ihre Zuordnung zur angeborenen Immunität rührt daher, dass sie nicht in der Lage sind, ein Gedächtnis aufzubauen: Wiederholte Applikationen desselben Antigens verändern ihre Immunantwort nicht [32]. Die Rolle der B-1-Zellen ist bisher nicht vollständig geklärt; vorgefunden werden sie vor allem in Brusthöhle und Bauchfell, was eine Funktion zum Schutz der Hohlräume nahelegt, während ihr eingeschränktes Repertoire eher für eine Rolle der angeborenen Immunantwort beim frühzeitigen Schutz vor Eindringlingen spricht. Bisher ist unklar, ob B-1-Zellen eine entscheidende Bedeutung bei der Immunabwehr haben oder

eher ein evolutionäres Relikt darstellen [32]. Ein grosser Teil der B-1-Zellen exprimiert ein breites Spektrum von selbstreaktiven Antikörpern [23]. Dennoch stammen bei unimmunisierten Mäusen grosse Teile des IgM-Pools von B-1-Zellen ab. Das Konzept der natürlichen Antikörper wurde bereits in den 50er-Jahren des 20. Jahrhunderts formuliert [33], doch hat bisher das Ansammeln von grossen Datenmengen nur in beschränktem Masse zur Klärung ihrer Existenz beigetragen.

### 1.2.4 Das Antikörperrepertoire

Das Antikörperrepertoire umfasst alle Immunglobuline, die ein Individuum bilden kann. Hier erscheint es sinnvoll, eine Unterteilung zu machen in

**potenzielles Repertoire** alle Antikörper, die ein Individuum bilden und in den Körper abgeben kann (d.h. ohne diejenigen Antikörper die bei den Selektions-Prüfstellen rausfallen)

und

**effektives Repertoire** alle Antikörper, die ein Individuum zu einem bestimmten Zeitpunkt zur Verfügung hat.

Die Produkte aller Reifungsprozesse der B-Zellen bilden ein effektives Repertoire von Antikörpern. Über die Dimension der enormen funktionalen Diversität dieses effektiven Repertoires eines Individuums ist bisher wenig bekannt [12]. Die theoretische Diversität des Repertoires der B-Zellen übersteigt  $10^{10}$  [9], während die beobachteten Komplexitäten von  $10^5$  bis  $10^7$  reichen [2], [14], [48]. Je früher die Messungen und Schätzungen stattfanden, desto grösser wurde das Repertoire geschätzt: Trepel [79] extrapolierte 1974 ein Repertoire von hunderten von Milliarden von Lymphozyten. Das Thema scheint nicht mehr sehr modern zu sein, im 21. Jahrhundert gibt es noch keine veröffentlichte Arbeit, die sich mit der Menge der unterschiedlichen Spezifitäten der Lymphozyten beschäftigt, obwohl die neuen Technologien Hochdurchsatzanalysen mit steigender Effizienz hervorbringen. Zurzeit gibt es keine verlässliche Abschätzung der Diversität – bzw. der Anzahl von unterschiedlichen und überschneidenden Spezifitäten der Antikörper – eines Repertoires.

### 1.2.5 Konzentrationen von Lymphozyten und Antikörpern

Die Anzahl der Antikörper-sezernierenden B-Zellen eines Individuums ist bisher nicht experimentell bestimmt. Es existieren nur wenige Publikationen, die Zahlen zu bestimmen versuchen. Der Kommentar von Liu [41] fasst zusammen, dass es  $10^8$  periphere B-Zellen in der Maus gibt. Opstelten u. Osmond [51] berechneten anhand immunologischer Färbungen die Proliferationsrate von pre-B-Zellen im Knochenmark und kamen auf  $0.5 \times 10^8$  Zellteilungen/Maus/Tag. Die Grösse des potenziellen Repertoires schätzt Liu [41] auf  $10^{11}$  in der Maus. Trepel [79] kommt in einer ähnlichen Schätzung auf 4 bis  $5 \times 10^{11}$  (400 bis 500 Milliarden) im jungen erwachsenen Menschen.

### 1.2.6 Spezifität, Diversität und Kreuzreaktivität von Antikörpern

Die *Spezifität* misst den Grad, zu welchem ein Sample-Molekül (hier: Antikörper) zwischen verschiedenen Probe-Bindungspartnern zu unterscheiden vermag [19]. Cohn u. Langman [14] definieren die Entstehung von Spezifität als den einzigen Selektionsdruck des Antikörpers, um zwischen selbst und nicht-selbst zu unterscheiden: »*The sole selection pressure driving antibody to be specific is the necessity to make the self-nonself discrimination*«[14][p. 36]. In dieser Definition wird Selbsterkennung gegenüber der Nicht-Selbsterkennung einbezogen, was in dieser Arbeit nicht besprochen wird, weil Selbsterkennung eine oft vorkommende Eigenschaft ist und krankhafte Symptome hervorrufen kann, wenn der selbsterkennende Antikörper eine falsche Effektorfunktion hat. Die Diskussion von selbst und nicht-selbst wird wohl eher schwierig und unpräzise bleiben, bis mehr und Genaueres über die Effektorfunktionen bekannt ist.

Ein Molekül, dessen Spezifität sich nur gegen ein einziges anderes Molekül richtet, ist per (überholter) Definition *monospezifisch*. Monospezifität (z. B.: Molekül *a* ist monospezifisch gegenüber Molekül *b*) bedeutet auch Molekül *a* erkennt Molekül *b* *x*-mal stärker als alle anderen (Getesteten). Glaubwürdige Monospezifität könnte nur deklariert werden, wenn ein Molekül *a* gegenüber **allen** möglichen Molekülen getestet worden wäre (was technisch nicht möglich ist), und mit Molekül *b* deutlich stärker als mit allen andern reagiert hätte.

Die *Diversität* hängt von der Anzahl verschiedener Antikörper und deren verschiedenen Spezifitäten gegenüber den Probe-Bindungspartnern ab. Allein aufgrund der grossen Zahl verschiedener Antikörper ist es unmöglich,

die Gesamtheit der möglichen Antikörper-Antigen-Interaktionen in einem bestimmten immunologischen Zustand zu erfassen. Die Gesamtheit der Interaktionen ist zudem einem ständigen Wechsel unterworfen, weil jeder Fremdkörper eine Immunantwort auszulösen vermag, wodurch das ganze Repertoire umgestaltet werden kann. Die Idee eines »Index der Variabilität«, der von der Anzahl verschiedener Klone eines Repertoires abhängt, wurde mittels quantitativer Western-Blot-Messungen in Brissac u. a. [12] untersucht und als linear abhängig postuliert.

Effekte wie die *Kreuzreaktivität*, mit der man die Fähigkeit von Sample-Molekülen (Antikörpern) bezeichnet, an unterschiedliche Probe-Bindungspartner (Antigene) mit unterschiedlichen Stärken zu binden [19, Kapitel 4], erhöhen die Komplexität solcher Interaktionen. Die Begriffe Kreuzreaktivität und Monospezifität werden in Abschnitt 1.4.1 besprochen. Die Kreuzreaktivität ist eine Eigenschaft aller biologischen Moleküle und insbesondere des Immunsystems, deren Auswirkung oft unterschätzt wird. In Anbetracht der Komplexität der verschiedenen Begriffe und der grossen Anzahl Mitwirkender (Antikörper) und der sogar noch grösseren Anzahl potenzieller Bindungspartner (alle existierenden und theoretisch herstellbaren Moleküle) ist es nicht erstaunlich, dass die meisten Studien, die das Antikörper-Bindungsverhalten untersuchen, isolierte Interaktionsprobleme zwischen kleinen Zahlen von Antikörpern und Antigenen betrachten. Ein erklärender Versuchsaufbau müsste die Bindungsreaktivitäten aller Antikörper eines Individuums zusammen in Betracht ziehen, zumindest wenn sich der Fokus der Forschung auf verschiedene Zustände des Immunsystems richtet oder wenn Krankheiten mit unbekannten Antigenen untersucht werden. Studien dieser Art sind selten; in den Abschnitten 1.4.2 und 1.6 folgen Übersichten.

### 1.2.7 Das Protekton

Der Begriff Protekton (engl. *protecton*) wurde von Langman u. Cohn [38] eingeführt. Die Protekton-Theorie beruht auf der Idee, dass die Immunsysteme von unterschiedlich massereichen Individuen (Maus, Mensch, Elefant) die jeweils gleiche minimale Schutzeinheit (*das Protekton*) benötigen, bzw. dass das Immunsystem von modularem Aufbau ist. Die Theorie geht davon aus, dass diese Einheit des minimalen Schutzes mit der Grössenzunahme eines Organismus vervielfacht wird. Aufgrund der zum Schutz eines Tieres benötigten Konzentration gehen Cohn u. Langman [14] davon aus, dass eine Protekton-einheit aus ungefähr  $10^7$  B-Zellen in einem Körpervolumenanteil von 1 ml Blut besteht. Von diesen  $10^7$  B-Zellen wiederum werden  $10^3$  als Plasmazellen

vorausgesetzt, die in der Lage sind, in 24 h 10 ng Antikörper zu sekretieren [14].

## 1.3 Immunantworten gegenüber Pathogenen

Die verschiedenen Pathogene, deren Immunantworten in dieser Arbeit untersucht wurden, werden hier kurz vorgestellt:

### 1.3.1 Nematoden und Nematodeninfektionen

Nematoden bilden den Stamm der Fadenwürmer (engl. *roundworms*). Zehntausende verschiedener Nematoden sind beschrieben, von denen ein Teil parasitär sowohl in Pflanzen als auch in Tieren (inklusive Menschen) lebt. Parasitäre Nematoden infizieren ihre Wirte unter anderem über Larven, die mit der Nahrung aufgenommen werden.

#### Infektionen mit *Heligmosomoides polygyrus*

Der Fadenwurm *Heligmosomoides polygyrus* (*H. polygyrus*) ist ein weitverbreiteter Parasit, der den Zwölffingerdarm von Mäusen und anderen Nagern bevölkert. In Mausmodellen dient er als Modellwurm zur Untersuchung von humanen Darmwürmern und der von ihnen hervorgerufenen Infektionen. Eine Übersicht der zellulären und molekularen Mechanismen der Interaktion mit dem Immunsystem von Helminth-Parasiten bietet die Publikation von Maizels u. Yazdanbakhsh [42].

### 1.3.2 Pertussis

Pertussis (Keuchhusten) wird durch das Bakterium *Bordetella pertussis* übertragen. Keuchhusten ist eine hochansteckende Krankheit und erlebt zurzeit trotz guter Impfquote ein weltweites Comeback [69]. Ausserdem ist Keuchhusten infolge von Komplikationen wie Lungen- oder Mittelohrentzündung keine ungefährliche Krankheit. Am meisten gefährdet sind Säuglinge, die infolge von Hustenanfällen an Atemnot sterben können. Bisher existiert keine Impfung mit Garantie auf Immunität, d.h., an Keuchhusten erkranken auch geimpfte Personen. Des Weiteren existiert kein diagnostisches Verfahren, um die Immunität gegenüber Pertussis zu testen.

## 1.4 Die Charakterisierung von monoklonalen Antikörpern und Serumantikörpern

### 1.4.1 Charakterisierung monoklonaler Antikörper

#### Monoklonaler Antikörper generell

Antikörper, die von einem oder von identischen Klonen von B-Zellen stammen, nennt man monoklonale Antikörper. Die Herstellung monoklonaler Antikörper erfolgt häufig mittels Hybridoma Zelllinien [35], einer Technologie der Fusionierung einzelner B-Zellen mit einer Myelomazelle (Neoplastische Plasmazelle), und inzwischen auch mittels gentechnischer Verfahren. Monoklonale Antikörper werden in serologischen Tests standardmässig als diagnostische Samples und auch als therapeutische Wirkstoffe benutzt. Monoklonale Antikörper sind entgegen veralteter Definitionen meistens nicht monospezifisch, was unter anderen die im folgenden aufgeführten Arbeiten der Charakterisierung monoklonaler Antikörper mittels grosser Peptid- und Proteinbibliotheken zeigen:

#### Die Charakterisierung monoklonaler Antikörper

Wichtige Arbeiten zur Charakterisierung monoklonaler Antikörper mittels kombinatorischer Bibliotheken stammen von Reineke u. a. [60], Michaud u. a. [45] und Predki u. a. [55]. Die erstgenannte Arbeit beruht ausschliesslich auf Bindungsmessungen monoklonaler Antikörper gegenüber synthetischen Peptiden. Mittels SPOT Synthese werden »systematische Bibliotheken« (Definition siehe Abschnitt 1.5.2) bestehend aus 5520 verschiedenen 15mer-Peptiden hergestellt. Die Autoren zeigen, dass die drei monoklonalen Antikörper CB4-1, CB/RS/13 und Tab2 bei Weitem nicht monospezifisch sind. Die Arbeit von Michaud u. a. [45] analysiert die Kreuzreaktivitäten von elf monoklonalen Antikörpern. Die hierfür benutzte Bibliothek hat eine »voreingenommene Zusammensetzung« (siehe ebenfalls Abschnitt 1.5.2); das ganze Proteom der Hefe wurde in ca. 5000 Proteinen auf Glasträger aufgetragen. Mittels Bindungsmessungen gegenüber diesen Glasträgern konnten die Autoren zeigen, dass die verschiedenen Antikörper in verschiedenem Ausmass mit der 5000er-Bibliothek kreuzreagieren. Und schliesslich liefert der Übersichtsartikel von Predki u. a. [55] eine Analyse der Kreuzreaktivitäten von Antikörpern gegenüber Protein-beladenen Glasträgern. Die Arbeit beschäftigte sich vor allem mit den technologischen Problemen und dem zu erwartenden therapeutischen Nutzen und zieht den Schluss, dass, selbst wenn das humane Proteom

in absehbarer Zeit auf einem Träger verfügbar sein sollte, die technologischen Probleme noch überwiegen werden.

### 1.4.2 Charakterisierung von Serumantikörper

Die ersten Bindungsstudien, die gesamte Antikörperpools im Blut von Mäusen und Menschen in Betracht zogen, wurden durch Inkubationen von Serumsamples auf mittels Western Blot aufgetrennten Proteingemischen aus Zellkultur-Extrakten durchgeführt [48], [22], [74]. Die Autoren waren in der Lage, anhand von Differenzen der Serumantikörper-Bindungsmuster verschiedene Krankheitszustände von gesunden Individuen zu unterscheiden. Die Unterscheidung der Gesundheitszustände war dabei allein aufgrund der Differenzen in der Erkennung der Western-Blot-Banden durch die Serumantikörper möglich. Eine Vielzahl von Publikationen zeigten, dass die optimierte Western-Blot-Technik in der Lage ist, sowohl verschiedene Inzucht-Mäusestämme zu klassifizieren [80] als auch Krankheiten zu diagnostizieren. Mittels Hauptkomponentenanalyse (engl. *PCA*, *Principal Component Analysis*), einer nicht überwachten (engl. *unsupervised* oder *unbiased*) Methode zur Dimensionsreduzierung der Daten, wurden Krankheiten anhand der Unterschiede in den Antikörperbindungsmustern analysiert und klassifiziert. Darunter:

1. die membranöse Glomerulonephritis (engl. *Membranous Glomerulonephritis (MGN)*) [71], eine autoimmunologische, chronische Nierenentzündung,
2. die Autoimmunthrombozytopenie (engl. *Idiopathic Thrombocytopenic Purpura (ITP)*) [76], eine Autoimmunkrankheit, die die Thrombozyten (Blutplättchen) betrifft und
3. die chronische lymphatische Leukämie (engl. *B-cell chronic lymphocytic leukemia (B-CLL)*) [75].

Klassifizierungen von Unterschieden, die nicht in den höchsten Varianzen enthalten waren, wurden mittels Linearer Diskriminantenanalyse (LDA), einer beaufsichtigten (engl. *supervised*) Lernmethode durchgeführt, wie z. B. in Stahl u. a. [72]. Das erwähnenswerte Ergebnis der Western-Blot-Resultate und ihrer Analysen ist die Tatsache, dass alle beschriebenen Beispiele linear klassifizierbar sind. Signifikanzanalysen, die zumindest bei LDA in hochdimensionalen Räumen mit wenig Datenpunkten angebracht wären, wurden

nicht beschrieben.

Neuere Arbeiten untersuchen murine Serumantikörper-Reaktivitäten mittels Hochdurchsatz-ELISA-Tests (engl. *Enzyme Linked Immunosorbent Assay (ELISA)*, dt. *Enzymgekoppelter Immunadsorptionstest (EIA)*) gegenüber 87 Proben [57] oder durch Inkubation der Sample-Seren auf protein- und peptidbeladenen Glasträgern [56]. Protein- und peptidbeladene Glasträger basieren auf einer neuen Hochdurchsatztechnologie, die ein sehr flexibles Bindungspartner-Bibliothek-Design ermöglicht, und somit eine ideale Methode zur Messung der Antikörper-Reaktivitäten darstellt. In der Arbeit von Quintana u. Cohen [56] bestand die Bibliothek aus verschiedenen, ausgewählten, selbstreaktiven Proteinen und Peptiden. Für die Merkmalsselektion, die zum Auffinden der Probe-Bindungspartner innerhalb der Bibliothek dient, die das grösste Differenzierungsvermögen zwischen gesunden, kranken und krankheitsanfälligen Mäuse-Individuen haben, verwendeten die Autoren den nichtparametrischen Wilcoxon-Vorzeichen-Rang-Test (engl. *Wilcoxon's matched pair test*). Die jeweils besten (ersten) zehn Prozent der mittels Wilcoxon-Test geordneten Merkmale ermöglichten nicht nur die Klassifizierung von gesunden gegenüber an Diabetes erkrankten Mäusen, sondern erlaubten darüber hinaus auch eine Klassifizierung von gesund bleibenden Mäusen gegenüber denen, die eine Anfälligkeit zur Diabeteserkrankung haben. Eine derartige Unterscheidung ist ein Meilenstein auf dem Weg zur Diagnose der Prädisposition von einzelnen Individuen gegenüber bestimmter Krankheiten.

Andere Arbeiten beschäftigen sich mit der Entwicklung von Krankheits-assoziierten Proteinträgern, um z. B. die Autoimmunkrankheit DCM (engl. *dilated cardiomyopathy*) zu diagnostizieren [28]. In der DCM-Arbeit wurden mehr als  $3.7 \times 10^4$  Proteine aus einer cDNA-Bibliothek von humanen, embryonalen Gehirnzellen hergestellt und auf Glasträger aufgetragen. Verschiedene Arbeiten verwenden den Ansatz die Proteine des Proteoms eines bekannten Organismus in Teilen oder als Ganzes als Probenbibliothek zu benutzen. Zur Vorhersage von SLE (Systemischer Lupus Erythematoses) wurden 30 Antigene aus dem Glomerulus-Umfeld exprimiert, auf Glasträger aufgetragen und mit den Seren verschiedener SLE-Maus-Modellen inkubiert. Mittels der Signalintensitäten der Bindungen konnten die Individuen entsprechend der verschiedenen Krankheitsverläufe gruppiert (engl. *clustered*) werden [40]. Abschnitt 1.6 bietet eine erweiterte Übersicht von Beispielen von Arbeiten mittels Hochdurchsatz-Analysen und der Vielfalt ihrer Standardisierungsversuche.



Zusammenfassend können wir festhalten, dass bei der gleichzeitigen Messung vieler Interaktionen von vielen einzelnen Molekülen mittels Hochdurchsatzverfahren grosse Mengen an Daten entstehen, die schwierig zu analysieren und interpretieren sind. Die Suche nach den wichtigsten Klassifikatoren in grossen Datenmengen, wie auch die Analyse der Signifikanz der Ergebnisse sind zumindest in der Immunologie bisher vernachlässigte Probleme.

## **1.5 Technologien der Bindungsdetektion mittels Hochdurchsatzstrategien**

### **1.5.1 Probemoleküle: Peptide eignen sich besser als Proteine**

Die Vorteile von Peptiden als Probemoleküle im Vergleich zu Proteinen sind in der folgenden Liste (pers. Kommunikation M. Schutkowski, JPT Peptide Technologies GmbH) zusammengefasst:

1. Die höhere Stabilität der Peptide erlaubt
  - (a) eine Versuchsdurchführung bei Raumtemperatur und
  - (b) eine bessere Lagerungsstabilität, welche Lagerungen über Jahre hinweg ermöglicht.
2. Der Vorteil der einfacheren, chemischen Synthese von Peptiden ermöglicht
  - (a) eine schnellere und billigere Synthese,
  - (b) die Umgehung von Faltungsproblemen,
  - (c) posttranslationale Modifikationen, die bestimm- und vergleichbar sind und
  - (d) eine höhere Vergleichbarkeit zwischen unterschiedlichen Produktionschargen (siehe Abschnitt 3.3.2 und 3.3.3).
3. Die Mengenvergleichbarkeit und -bestimmung ist bei Peptiden viel einfacher.

Generell gilt: Synthetische Peptide sind reproduzierbarer und dadurch geeigneter für alle Anwendungen, die sich mit diagnostischen Fragen beschäftigen. Aus diesen Gründen beziehen sich die folgenden Absätze nur noch auf Peptide.

## 1.5.2 Peptidbibliotheken

Peptidbibliotheken sind in erster Linie Zusammenstellungen verschiedener Peptide. Die Peptide können dabei aus natürlichen Quellen, wie z. B. aus Zell- und Gewebeextrakten oder aus Expressionssystemen stammen, oder chemisch synthetisiert werden. Je nach Herkunft und Anordnung sind die Sequenzen der einzelnen Peptide bekannt und/oder räumlich adressierbar. Des Weiteren sind Peptide, die aus eukaryontischen Zellen stammen, posttranslational modifiziert. Die posttranslationalen Modifikationen sind nur schwer bestimmbar (mit Massenspektrometrie) und nicht reproduzierbar. Eine Tatsache, die oft ignoriert wird, aber von zunehmender Wichtigkeit ist, da sich die Hinweise mehren, dass posttranslationale Modifikationen das Faltungs- und Bindungsverhalten von Proteinen stark beeinflussen.

### Herstellungsverfahren kombinatorischer Peptidbibliotheken

Kombinatorische Peptidbibliotheken können biologischen Ursprungs sein, wie z. B. die Phagen-Display-Bibliotheken (engl. *phage display libraries*), die 1988 von Parmley u. Smith [52] eingeführt wurden. Bei der Suche nach Artikeln (seit 1995) unter den *MeSH*-Begriffen »epitope mapping« und »peptide library« in *Pubmed* finden sich 265 Treffer, von denen sich 147, d.h. mehr als die Hälfte (55%), auf den Begriff »phage display« beziehen. Biologische Bibliotheken können sowohl aus *in vivo* als auch aus *in vitro* Translationssystemen stammen. Nebst Phagen-Display basieren weitere Beispiele von *in vivo* Expressions- oder Translationssystemen auf Plasmiden oder Hefezellen. Ein Beispiel zur Herstellung von *in vitro* Peptiden mittels biologischer Hilfe ist die Technologie des Ribosomen-, bzw. des Polysomen-Displays [43]. Der Nachteil biologischer Bibliotheken ist, dass sie nur mit grossem Aufwand räumlich adressierbar sind, und die gefundenen Probe-Bindungspartner immer durch Klonierung, Vermehrung und/oder Sequenzierung bestimmt werden müssen. Positiv hervorzuheben ist jedoch, dass biologische Bibliotheken bis zu  $10^9$  verschiedene Peptide präsentieren können, deren Länge von bis zu 30 Aminosäuren die Länge von synthetisch hergestellten Peptiden mit ausreichender Reinheit bei Weitem übersteigt.

Die erste synthetische, räumlich adressierbare Bibliothek wurde mittels Anhaft-Technologie (engl. *pin-up*) in 96-Napf-Platten von Geysen u. a. [20] synthetisiert. Dabei konnten bereits hunderte verschiedener Peptide parallel synthetisiert und untersucht werden. Anordnungen verschiedener Peptidsequenzen, die eindeutig räumlich zugeordnet werden können haben viele Vorteile gegenüber Peptidmischungen. Die SPOT Synthese [18] (siehe auch Ab-

schnitt 3.3.1) erlaubt Experimente zur Suche nach Probe-Bindungspartnern, die direkt auf dem Syntheseträgermaterial durchgeführt werden können. Neue Ansätze wie Peptidbibliotheken auf Glasträgern ermöglichen mehr Peptidspots pro Fläche. Darüber hinaus bieten Glasträger den Vorteil, dass man die Peptide nur einmal synthetisieren muss und danach grosse Mengen von Trägern mit identischen Bibliotheken gedruckt und inkubiert werden können. Detailliertere Vergleiche zwischen Zellulose und Glas als Inkubationsträger folgen in Abschnitt 8.5.

### **Zusammensetzungen der Peptidbibliotheken**

Peptidbibliotheken lassen sich in »voreingenommene« und »systematische Zusammensetzungen« unterteilen. Für die vorliegende Arbeit wurden ausschliesslich systematische (unvoreingenommene) Zusammensetzungen von Proben benutzt. Darin unterscheidet sie sich von den meisten anderen Arbeiten. Deshalb werden die Begriffe systematisch und voreingenommen hier genauer definiert:

**Voreingenommene** Zusammensetzungen von Peptiden bezeichnen Bibliotheken, die auf das Ziel des Versuchs abgestimmt sind. Voreingenommene Bibliotheken sind z. B. solche, die nur Peptide aus dem Proteom eines Virus enthalten, um Serumsamples auf Reaktivität gegenüber diesen viralen Peptiden zu überprüfen, oder Bibliotheken von »Selbstantigenen« zum Auffinden von autoreaktiven Antikörpern.

**Systematische (unvoreingenommene)** Zusammensetzungen sind per Definition nicht auf ein Problem ausgelegt, d.h. die Peptidsequenzen sind rein zufällig oder haben höchstens eine Ausrichtung, welche z. B. die Diversität der Proben beeinflusst, aber unabhängig von der untersuchten Fragestellung ist. Systematische Bibliotheken sind unabhängig vom Verlauf oder Grund einer Immunantwort.

### **1.5.3 Peptidsynthese: Die SPOT Synthese**

#### **SPOT Synthese**

Die SPOT Synthese wurde 1988 von Frank u. Döring [18] beschrieben und bietet gegenüber früheren Technologien der Peptidsynthese den Vorteil, dass auf räumlich getrennten, aber einzeln adressierbaren Stellen einer Zellulosemembran verschiedene Peptide parallel synthetisiert werden können.

Eine Zusammenstellung wichtiger Fakten (pers. Kommunikation U. Reineke, JPT Peptide Technologies GmbH) folgt hier:

**Spotgrösse** – Je nachdem, ob die Peptide direkt auf der Zellulose inkubiert oder ausgelöst und weiterverarbeitet werden, synthetisiert man »kleinere« oder »grössere« Spots. Kleinere Spots haben Durchmesser von ca. 1 mm und grössere von ca. 9 mm.

**Reinheit** – Die Reinheit der Peptidsynthese bei vollautomatischer Synthese (siehe Abschnitt 3.3) beträgt 60% für 15mere und nur noch 10% für 30mere, wobei grössere Spots eine höhere Synthesereinheit aufweisen.

**Zugänglichkeit** – 10 bis 20% der Peptide auf einer Zellulosemembran sind für Antikörper zugänglich.

**Peptidbeladung** – Beim Vergleich der Beladungen der Zellulosemembran (siehe Abschnitt 3.3) finden sich ca. 10% Differenz zwischen der Beladung am Anfang und am Ende der Peptidsynthesen. Die Zahl der abgebrochenen Peptidsynthesen ist jedoch kleiner als 10%. Was mit den Peptiden geschieht, deren Synthese weder abgebrochen noch fortgeführt wird, ist unklar.

## 1.5.4 Trägermaterialien: Makro- und Mikroarrays

### Makroarrays: Zellulosemembranen

Inkubationen auf Peptidbibliotheken auf Zellulosemembranen gehören in die Kategorie der Makroarray-Versuche, weil die Ausdehnung der einzelnen Spots eher gross ist.

### Mikroarray: Peptidbeladene Glasträger

Die »PepStar«-Technologie ist eine eingetragene Marke von JPT Peptide Technologies GmbH, Berlin. JPT bezeichnet »PepStar« als eine Peptid-Mikroarray-Plattform zur Herstellung von massgeschneiderten, serienmässig hergestellten Peptid-Mikroarrays auf Glasträgern mit hohem Peptidgehalt (engl. *high content*). Die Peptide werden kundenspezifisch mittels SPOT Synthese massgefertigt (d.h. synthetisiert), aus der Zellulose ausgelöst und in gelöstem Zustand mittels Kontakt- oder Nicht-Kontakt-Verfahren auf Glasträger aufgetragen. JPT ist eine der wenigen Firmen, die solche peptidbeladenen Glasträger in kurzer Zeit in grossen Mengen herstellen kann.

Unter Verwendung von peptidbeladene Mikroarrays wurden in den letzten Jahren zuverlässige Methoden zur Messung von Protein-Interaktionen in gross angelegten Studien mit hohem Durchsatz entwickelt. Eine Übersicht

der wichtigsten Arbeiten bieten Abschnitt 1.6 und die zwei Übersichtsartikel von Andresen u. a. [3] und Zhu u. a. [87].

Die wichtigsten Unterschiede von Zellulose (Makro) und Glas (Mikro) als Trägermaterial sowie ihre Vor- und Nachteile werden in Abschnitt 8.5 detaillierter diskutiert.

	<b>Zellulosemembran</b>	<b>Glasträger</b>
Trägermaterial	Zellulose	Glas
Hersteller	Whatman	Genetix XL ErieScientific
Peptid Immobilisierungs-Richtung	C-Terminus	N-Terminus
Oberflächenchemie	amin/amid	aldehyd/oxim
Peptidkonzentration	250000 fmol/mm <sup>2</sup>	15 fmol/mm <sup>2</sup>
Detektion (Markierung 2ter Ak)	Chemolumineszenz (POD)	Fluoreszenz (Alexa Fluor)
Inkubationsvolumen pro 100 Spots	4000 $\mu$ L auf 200 $\mu$ L reduzierbar	12 $\mu$ L

**Tabelle 1.1:** Der Vergleich von Makroarrays (auf Zellulosemembranen) und Mikroarrays (auf Glasträger).

## 1.6 Serumantikörper-Reaktivitäts-Profile auf Mikroarrays in der Literatur

Die existierenden Publikationen bezüglich Mikroarrays und Serumantikörper-Bindungsmessungen widersprechen der Aussage, dass Peptide als Probemoleküle in Bindungspartnerbibliotheken geeignetere sind als Proteine (Abschnitt 1.5.1). Die meisten Arbeiten benutzen überwiegend Proteine als Probemoleküle. Diese Arbeiten, von denen es nur ein paar Dutzend gibt, sind alle in den letzten Jahren entstanden. Die Datenvorverarbeitung und -normierung wurde meistens direkt von der DNA-Mikroarray-Datenanalyse übernommen, was aufgrund der Unterschiede in den Technologien (Antikörper-Protein-Bindungen sind stabiler als DNA-Hybridisierungen) ein riskantes Vorgehen ist. Eine Auflistung der Normierungsvorgehen der wichtigsten Arbeiten folgt hier:

- Die Signale des lokalen Trägerhintergrundes wurden von den SI der Peptidspots abgezogen (beispielsweise von Horn u. a. [28], Neuman de

Vegvar u. a. [81], Quintana u. Cohen [56], Robinson u. a. [63] und Tong u. a. [78]).

- Die Signale wurden auf positive Signale, wie sekundäre-Antikörper-Bindung oder Proteinkonzentration normiert (bei Horn u. a. [28], Hiller u. a. [24], Hueber u. a. [29], Li u. a. [40], Robinson u. a. [64], Shreffler u. a. [68]).
- Aufgrund der Streuung der Signale wurden Mittelwerte oder Mediane von 2 – 8 Replikaten identischer Peptidspots gebildet (bei Chatterjee u. a. [13], Hiller u. a. [24], Horn u. a. [28], Hueber u. a. [29], Neuman de Vegvar u. a. [81], Quintana u. Cohen [56], Robinson u. a. [63], Robinson u. a. [64], Shreffler u. a. [68], Tong u. a. [78]).

Die Serienverdünnungen reichten von 1 : 5 bis 1 : 300, wobei mehrheitlich Verdünnungen im  $10^{-2}$ -Bereich angewandt wurden. Die immobilisierten Probenbibliotheken waren alle voreingenommen (Abschnitt 1.5.2) und bestanden mehrheitlich aus Proteinen, die teilweise mit Peptiden ergänzt wurden. Als weiteren vielversprechenden Probenbibliothek-Ansatz seien hier noch Mikroarrays mit verschiedenen Zuckerringen als Problemoleküle aufgeführt (von Wang u. a. [83]).

## 1.7 Dimensionsreduktion

Wenig Datenpunkte (Samples) in einem hochdimensionalen Raum (d.h. viele Proben oder Variablen oder Merkmale) sind schwierig zu analysieren, da Klassifizierungsverfahren von wenigen Samples mittels vieler Proben zu Überanpassung (engl. *overfitting*) tendieren. Eine Reduktion der Dimension ist daher oft hilfreich. Eine Dimensionsreduktion dient der Reduktion der Anzahl der gemessenen Variablen (d.h. Merkmalen oder Proben) und erfolgt mittels unvoreingenommener Dimensionsextraktion oder mit ausgefeilteren Methoden der Merkmalsselektion.

### 1.7.1 Unvoreingenommene Dimensionsextraktion

Unvoreingenommene Dimensionsextraktion bezeichnet die Reduktion der Anzahl Dimensionen eines Datensatzes ohne auf die Wichtigkeit der einzelnen Merkmale bezüglich der Fragestellung Rücksicht zu nehmen. Unvoreingenommene Dimensionsextraktion erfolgt beispielsweise durch Elimination zufällig ausgewählter Proben. Ein anderes Beispiel der unvoreingenommenen Dimensionsextraktion ist die Hauptkomponentenanalyse (PCA). PCA sucht

die Richtung der maximalen Varianz und transformiert eine Datenmatrix in ein neues Koordinatensystem, in dem die erste Koordinate die höchste Varianz, die zweite Koordinate die zweithöchste Varianz usw. enthält. Durch die Wahl der Zahl der Koordinaten, die Hauptkomponenten genannt werden, kann dann eine beliebige Zahl von Dimensionen gewählt werden, die bei der ersten beginnend grosse Teile der Information des Datensatzes enthalten.

### 1.7.2 Merkmalsselektion

Merkmalsselektion dient ebenfalls der Reduktion der Anzahl der Dimensionen eines Datensatzes. Darüber hinaus wertet die Selektion die Merkmale nach einer bestimmten Vorschrift. Es existieren viele, je nach Problem mehr oder weniger geeignete Methoden zur Merkmalsselektion. Einige Methoden der Merkmalsselektion, wie z. B. der Wilcoxon-Vorzeichen-Rang-Test, dienen nur zum Erstellen einer Reihenfolge aller Merkmale, während andere Selektionsmethoden auch die Gruppen der Merkmale und die Ergänzung der Merkmale untereinander bewerten. Die Methoden der Merkmalsselektion lassen sich gemäss Saeys u. a. [66] in »Wrapper«-, »Filter«- und »Eingebettete«-Methoden einteilen.

## 1.8 Mathematische Modellierung von Wechselwirkungen zwischen Antikörpern und Antigenen

### 1.8.1 Das Konzept des »Shape Space«

Das Konzept »Shape Space« geht auf Perelson u. Oster [54] zurück. Das Modell beruht auf der Idee, dass alle Antikörper und Antigene als Punkte in einem Raum (dem *Shape Space*) betrachtet werden, und die Distanzen zwischen ihnen ein Mass ihrer Affinitäten abbilden. Die den Distanzen entsprechenden Affinitäten zwischen Antikörper-Rezeptor und Antigen-Epitop lassen sich verschiedenen Interaktionsklassen wie elektrostatische Ladung, Wasserstoffbindung, Van-der-Waals-Interaktionen und anderen elektrostatischen, chemischen und geometrischen Eigenschaften zuordnen. Durch die Auflistung dieser Interaktionsklassen eines jeden Moleküls in einem Vektor  $L$  kann jedes Molekül als Punkt in einem  $L$ -dimensionalen Raum dargestellt werden. Die verschiedenen Interaktionsklassen korrelieren teilweise, während andere komplementär zueinander sind; die Klassen überschneiden sich aber auch, was zu einer Dimensionsreduktion führt. Eine anschauliche Dimen-

sionsreduktion von 26 physikochemischen Eigenschaften aller Aminosäuren (die wir zu diesem Zweck als Unterraum der Moleküle betrachten können) auf fünf repräsentative Deskriptoren ist auch in Sandberg u. a. [67] publiziert (unabhängig von »Shape Space«). Die Publikation ist nahezu hundert Mal referenziert, und zeigt den Sinn der Dimensionsreduktion anhand quantitativer Messgrößen.

Laut Shape-Space-Theorie ist die Gesamtheit aller Antikörper-Antigen-Interaktionen ebenfalls in einem Raum von fünf bis zehn Dimensionen repräsentierbar. Ein solcher Raum dient nicht nur der Erklärung des komplexen Netzwerks der Antikörpererkennung im Immunsystem, sondern im Idealfall auch zur Vorhersage von Bindungsaffinitäten. Die Erhebung der Messdaten der Bindungs-Interaktionen, das Koordinatensystem, die Metrik und die Transformationsfunktion der Datenpunkte des Shape Space sind bisher ungelöste Probleme. Ein Shape Space, der auf Affinitätsmatrizen beruht, wurde von B-Rao u. Stewart [6] diskutiert, wobei mittels multidimensionaler Similaritätsstrukturanalyse (SSA, engl. *similarity structure analysis*), die eine frei wählbare Metrik und Transformationsfunktion erlaubt, verschiedene Datensätze grafisch dargestellt wurden. Dieser Ansatz ist jedoch nur in der Lage, die Ähnlichkeiten der einzelnen Moleküle, nicht aber deren Komplementarität abzubilden. Die Repräsentation von Lapedes u. Farber [39] von Hämagglutinin-Hemmungs-Messungen von 14 Antiseren gegenüber 19 Influenzaviren hatte ähnliche Probleme. Der Datensatz konnte jedoch mittels multidimensionaler Skalierung (MDS) erfolgreich auf einen vier- bis fünfdimensionalen Datenraum reduziert werden. Messpunkte wurden hierfür nur nach Grösse geordnet, ohne Berücksichtigung der absoluten Werte. Dabei wurden die Seren als Antikörpermischungen betrachtet. Wie sehr Affinitäts- oder Inhibitionsmessungen und deren mangelnde Exaktheit zur Charakterisierung von Antikörper- und Antigeninteraktionen reichen, ist bis heute nicht abschliessend untersucht. Des Weiteren kann man nicht ausschliessen, dass die gemessenen Interaktionen unterschiedliche Epitope betreffen, was bei unendlich grossen Messreihen vernachlässigt werden kann, in Versuchen mit kleineren Datenmengen aber sehr wohl eine Rolle spielt. Die Parameter des »Shape Space« und der Kreuzreaktivität sowie der Einfluss der Metrik auf die Dimension des Raumes wurden auch von Smith u. a. [70] diskutiert. Die Definitionen eines »Shape Space« hängen von den Spezifitäten und Kreuzreaktivitäten der Moleküle ab. Sie sind auch die bestimmenden Parameter, die das Postulat der Metrik des Shape Spaces von Perelson u. Oster [54] bestätigen können. Die Frage nach der Wahl der Methode der Dimensionsreduktion anhand von Bindungsmessungen und die Frage, ob Dimensionsreduktionen zu Bindungs-Vorhersagen führen können, werden sich beantworten



lassen, wenn mehr passende Messdaten zur Verfügung stehen. In der vorliegenden Arbeit wird ein Ansatz zur Dimensionsreduktion von Antikörper-Reaktivitäts-Profilen gezeigt, der sich für diagnostische Differenzierungen als geeignet herausgestellt hat.



# Kapitel 2

## Zielsetzung

Diese Arbeit untersucht die Unterschiede der Bindungseigenschaften von Serumantikörpern verschiedener Gruppen von Individuen mittels der Analyse von Serumantikörper-Reaktivitäts-Profilen. Das erfordert die Entwicklung und Evaluation geeigneter Technologien zur Messung der Bindungen und die Wahl geeigneter Methoden zur Datenanalyse.

Die Grundidee basiert auf der These, dass Immunreaktionen auf humoraler Ebene in ihrer Komplexität nur zu verstehen sind, wenn wir eine Technologie finden, mit der die Bindungsreaktivitäten des Antikörperpools in ihrer Gesamtheit betrachtet werden können. Dafür müssen die Bindungspartner der Antikörper einen »sinnvollen« Raum abdecken (d.h. eine geeignete Probenbibliothek präsentieren). Aus unserer Sicht ist es zurzeit und wohl auch in Zukunft nicht möglich, eine Bibliothek von Bindungspartnern herzustellen, mit der alle potenziellen Reaktivitäten und Kreuzreaktivitäten eines ganzen Antikörperrepertoires gemessen werden können. Wir wollen präsentieren, dass die Messung einer Teilmenge der Bindungsreaktivitäten jedoch ausreichen kann, um Antikörperrepertoires verschiedener Gruppen zu unterscheiden.

In der Literatur wurde mehrfach gezeigt, dass die Differenzierung verschiedener Krankheiten aufgrund der Messung von Serumantikörper-Reaktivitäts-Profilen möglich ist. Vor allem die Western-Blot-Publikationen, die aufgrund ihrer Probemoleküle (denaturierten Proteinen aus Gewebeextrakten) nur schlecht reproduzierbar sind, haben bewiesen, dass der experimentelle Ansatz zu klassifizierbaren Ergebnissen führt. Die erwähnten Arbeiten betrachteten ihre Ergebnisse allerdings nicht unter diagnostischen Gesichtspunkten. Sie klassifizierten zwar verschiedene Gruppen von kranken und gesunden Individuen aufgrund der Unterschiede in den Bindungen der

Serumantikörper-Reaktivitäts-Profile, versuchten jedoch nicht, einzelne Individuen vorherzusagen. Auch eine genaue Untersuchung der Technologie wurde nicht vorgenommen. Dafür zeigen die in den 1990er-Jahren publizierten Ergebnisse der Western-Blot-Experimente [48], [22], [74], dass sich Serumantikörper-Reaktivitäts-Profile unterscheiden. Die Streuung der Reaktivitäten gesunder Individuen im Vergleich zur Varianz zwischen »gesund« und »krank« ist darüber hinaus klein genug, um gesunde von kranken Individuen klassifizieren zu können. Ein weiterer Hinweis dafür, dass die Varianz von humanen Antikörper-Bindungsmustern nicht sehr gross ist, ist eben die Tatsache, dass die Anzahl der Polymorphismen der B-Zell-Rezeptor-Gensegment-Allele, die für humane Antikörper kodieren, erstaunlich klein ist, und auch die gewählten Rekombinationen den Rahmen des potenziell Möglichen nicht annähernd ausschöpfen, wie von Ignatovich u. a. [30] beschrieben.

An diesen Stellen setzen wir mit einem geeigneteren Verfahren an, um die Reproduzierbarkeit und Analyse der Serumantikörper-Reaktivitäts-Profile zu verbessern und in Richtung Diagnostik weiterzuentwickeln. Durch die Verwendung von Peptidarray-Hochdurchsatztechnologie erreichen wir eine höhere Reproduzierbarkeit der gemessenen Bindungen. Des Weiteren verwenden wir Merkmalsselektion und untersuchen die Eigenschaften der unterscheidenden Merkmale. Die Strategie unserer Merkmalsselektion unter Beachtung der Gruppenzusammensetzung ist eine bisher nicht publizierte Vorgehensweise.

Die mittelfristige Weiterentwicklung dieser Arbeit zielt darauf ab, mit den beschriebenen Verfahren eine nicht-invasive Diagnostikanwendung zu entwickeln, welche ohne a-priori-Wissen über Antigene auskommt. Für das grosse Potenzial eines anwendungsfreundlichen Diagnostikverfahrens sprechen die Sensitivitäten und Spezifitäten, die an diejenigen von kommerziellen Tests heranreichen, sowie die Aussagekraft der Merkmalsselektion, die eine standardisierte Auswertung ohne Normierungsaufwand ermöglicht. Es ist zu erwarten, dass die experimentelle Plattform (d.h. die synthetischen Peptide) in Kombination mit der differenzierten Datenanalyse (d.h. Merkmalsreduktion) den Transfer der Ergebnisse auf eine preiswerte, einfach zu handhabende Diagnostik-Plattform erlauben wird.

# Kapitel 3

## Material und Methoden

### 3.1 Mäuse und Serengewinnung

#### 3.1.1 Mäuse

Die BALB/c- und C57BL/6J-Mäuse männlichen Geschlechts wurden unter keimfreien Bedingungen gehalten und im Alter zwischen 6 und 10 Wochen geblutet. Parasiteninfizierte Mäuse wurden von der Arbeitsgruppe *Molekulare Parasitologie, HU Berlin* zum Zweck der eigenen Forschung infiziert und geblutet. Die Seren der gesunden und infizierten Mäuse wurden im Rahmen von Kooperationen an uns abgegeben. Die Mäuseseren stammen von unterschiedlichen Quellen:

- *Charles River Deutschland GmbH, Sulzfeld,*
- Arbeitsgruppen *C. Berek* und *S. Fillatreau, Deutsches Rheumaforschungszentrums (DRFZ), Berlin,*
- Arbeitsgruppe *S. Hartmann, Molekulare Parasitologie, HU Berlin.*

#### 3.1.2 *H.-polygyrus*-Infektionen

Die untersuchten BALB/c-Mäuse wurden mit dem im Darm ansässigen Fadenwurm (engl. *intestinal nematode*) *Heligmosomoides polygyrus*-infiziert. Die Mäuse wurden oral mit 150 infektiösen *H.-polygyrus*-Larven (L3) inokuliert. Erfolgreiche Infektionen wurden sowohl durch fäkale Eierzählung (Eierausscheidung beginnt zehn Tage nach Infektion), wie auch durch Rückgewinnung von adulten Würmern aus dem Dünndarm nach Tötung der Tiere überprüft.

### 3.1.3 Serengewinnung

Das Blut von narkotisierten Mäusen wurde direkt aus dem Herzen oder aus den Augenhöhlen abgenommen. Nicht-infizierte Mäuse wurden vollständig ausgeblutet, während von den *H.-polygyrus*-infizierten Mäusen zu verschiedenen Zeitpunkten nach der Infektion Blutsamples abgenommen wurden. Dazu wurden die Mäuse mit Ketamin/Rompur betäubt. Den betäubten Mäusen wurde dann mit Mikro-Hämatokrit-Kapillaren aus dem Augenhintergrund Blut entnommen. Die entnommene Blutmenge pro Individuum pro Zeitpunkt nach Infektion entsprach ca. 200  $\mu\text{l}$ . Nach der Abnahme wurde das Blut bei Raumtemperatur zur Gerinnung stehen gelassen, die Zellen abzentrifugiert und der Überstand (das Serum) aliquotiert und bei  $-20^{\circ}\text{C}$  eingefroren.

## 3.2 Peptidbibliotheken

### 3.2.1 Design von Peptidbibliotheken mit zufälligen Aminosäuresequenzen

Die Peptidsequenzen, die auf den Zellulosemembranen 15 Aminosäuren (AS) und auf den Glasträgern 14 AS lang sind, wurden mittels eines selbstgeschriebenen Zufallsgenerator-Skripts in *Perl* erzeugt. Hierbei wurden die einzelnen Aminosäuren (inklusive Cystein) nur mit der Frequenz in die Sequenzen eingesetzt, mit der sie auf den Oberflächen von Proteinen vorkommen. Mit Oberflächen von Proteinen ist der Teil der Proteine gemeint, der zugänglich für Flüssigkeiten ist (flüssigkeitszugängliche Oberfläche, bzw. *sas* von engl. *solvent accessible surface*). Die Frequenz der einzelnen Aminosäuren auf der *sas* wurde von *Kristian Rother, Mathematische Systembiologie, Charité Berlin* berechnet. Die Frequenzen wurde mittels eines 1.4-Å-Probemoleküls bestimmt, welches gegenüber  $1.3 \times 10^6$  Atomen in 3D-Proteinstrukturen getestet wurde. Jedes Atom, das von dem kleinen Probemolekül berührt werden konnte, wurde als zur Oberfläche zugehörig definiert. Die Atome wurden nach Aminosäuren zusammengefasst, unabhängig davon, ob sie stark oder gering exponiert waren. Die Werte in Tabelle 3.1 geben den Prozentsatz der molekularen Oberflächen von Proteinen an, die von Atomen der Aminosäure *k* eingenommen werden. Zum Vergleich mit dem totalen prozentualen Vorkommen der einzelnen Aminosäuren sind auch die Zahlen der innen liegenden Atome und die Gesamtzahlen der Atome pro Aminosäure auf die gleiche Weise ausgezählt worden.

	Oberfläche		Innen		Gesamt	
Aminosäure	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Ala	3832	5.5	3733	5.9	7565	5.7
Cys	806	1.2	1230	1.9	2036	1.5
Asp	5773	8.3	2907	4.6	8680	6.5
Glu	5667	8.1	2366	3.7	8033	6.0
Phe	2538	3.6	5261	8.3	7799	5.8
Gly	3715	5.3	2217	3.5	5932	4.4
His	1859	2.7	1711	2.7	3570	2.7
Ile	2140	3.1	4847	7.6	6987	5.2
Lys	6297	9.0	2356	3.7	8653	6.5
Leu	3433	4.9	6831	10.7	10264	7.7
Met	1034	1.5	1613	2.5	2647	2.0
Asn	4289	6.2	2675	4.2	6964	5.2
Pro	3359	4.8	2206	3.5	5565	4.2
Gln	3510	5.0	2023	3.2	5533	4.1
Arg	5109	7.3	2829	4.4	7938	6.0
Ser	4401	6.3	3003	4.7	7404	5.5
Thr	4324	6.2	3527	5.5	7851	5.9
Val	2864	4.1	5478	8.6	8342	6.3
Trp	1347	1.9	2489	3.9	3836	2.9
Tyr	3375	4.8	4434	7.0	7809	5.9
Gesamt	69672	100	63736	100	133408	100

**Tabelle 3.1:** Verteilung der Aminosäuren auf der Oberfläche von Proteinen. Die Häufigkeiten der einzelnen AS werden mittels eines Probemoleküls berechnet. *N* steht für die Anzahl der ausgezählten Atome.

### 3.3 Peptidsynthese und Herstellung der Peptidarrays

Die Peptide der Zellulosemembranen und der Glasträger wurden alle mittels SPOT Synthese auf Zellulose synthetisiert.

#### 3.3.1 SPOT Synthese

Die Synthese der Peptide erfolgte durch manuelle (halbautomatische) oder vollautomatische SPOT Synthese (Multi pep Roboter (INTAVIS Bioanalytical Instruments AG, Köln, Deutschland)) auf Zellulose. Trägermaterial der Synthese war eine mit Fmoc- $\beta$ -Ala bzw. bei Verwendung von CAPE-Mem-

branen (CAPE, von engl. *CelluloseAminoPropylEther*) mit Aminopropylether Amino-funktionalisierte Zellulosemembran (Whatman 50, Whatman, Maidstone, UK). Nach Abspaltung der Fmoc-Schutzgruppe durch 20% Piperidin wurden die Positionen der einzelnen Spots durch eine weitere Funktionalisierung mit  $\beta$ -Alanin durch positionsgenaues Spotten einer 0.3 M Lösung von Fmoc- $\beta$ -Alanin-OPfp und N-acetyl- $\beta$ -Alanin-OPfp (1:1) in DMSO (Dimethylsulfoxid) definiert. Anschliessend wurden freie Aminofunktionen acetyliert und die Beladung der Zellulosemembran durch photometrische Fmoc-Bestimmung berechnet. Die Synthese der Peptide erfolgte dann Schritt für Schritt vom C- zum N-Terminus nach Standard-Syntheseprotokoll, parallel auf den vordefinierten Spots. Eine weitere Fmoc Abspaltung diente der quantitativen Analyse der synthetisierten Peptide. Nach dem letzten Kopplungsschritt wurden die Seitenkettenschutzgruppen mit 90% TFA (Trifluoroessigsäure, engl. *trifluoro acetic acid*) abgespalten. Detaillierte Informationen zur SPOT Synthese finden sich in [17], [36], [86].

### 3.3.2 Herstellung von Peptidarrays auf Zellulosemembranen (Makroarrays)

Die Peptidarrays auf Zellulosemembranen wurden in unterschiedlichen Chargen synthetisiert. Chargen von Zellulosemembranen sind definiert als Zellulosemembranen, die zur gleichen Zeit auf dem gleichen Stück Zellulose, das in verschiedene Unterarrays (engl. *subarray*) unterteilt worden ist, synthetisiert worden sind. Alle Zellulosemembranen aus einer Charge wurden jeweils mit identischen Sequenzen an identischer  $x,y$ -Position der einzelnen Subarrays synthetisiert, d.h. jede Membran aus derselben Charge hatte die gleichen Peptide am gleichen Ort angeordnet, und wurden danach mit einem unterschiedlichen Serum inkubiert.

Aufgrund der Grösse der Auflagefläche unter dem Pipettier-Roboter und in der Waschwanne ist eine maximale Grösse von  $19 \times 29$  cm der Zellulosemembran vorgegeben und somit auch eine maximale Anzahl Peptidspots bzw. eine maximale Anzahl identischer Unterarrays (engl. *subarrays*) für eine gegebene Anzahl verschiedener zu synthetisierender Peptide. Kleinere Membranen mit rund 170 Peptidspots passen in 20facher Wiederholung auf eine Charge, während grosse Membranen mit rund 600 Peptidspots nur sechsmal Platz haben.

Eine weitere Differenzierung der Herstellungsverfahren der Zellulosemem-



branen besteht zwischen manueller (bzw. halbautomatischer) und automatischer Herstellung. Im manuellen Verfahren wurde die Zellulosemembran nach jeder Aufpipettier-Runde aller den Sequenzvorgaben entsprechenden Aminosäuren vom SPOT-Roboter entfernt und in einer separaten Waschwanne entsprechend dem Protokoll manuell gewaschen. Vor dem nächsten Pipettierschritt wurde die Zellulosemembran neu auf dem Pipettierroboter befestigt und korrekt platziert. Im automatischen Verfahren fanden Aufpipettieren der Aminosäuren und dazwischenliegende Waschschrte in der gleichen Wanne statt.

Die mittels SPOT Synthese hergestellten Peptide wurden nach Abspaltung der Seitenschutzgruppen direkt auf der Zellulose, auf der sie synthetisiert wurden, weiterverwendet (inkubiert).

### **3.3.3 Herstellung von Peptidarrays auf Glasträgern (Mikroarrays)**

Die Herstellung der Mikroarrays und die Synthese der aufzutragenden Peptide erfolgten bei JPT Peptide Technologies GmbH, Berlin. Hier eine kurze Übersicht der Herstellung: Die durch SPOT Synthese hergestellten Peptide wurden mittels wässriger Triethylamin-Lösung aus der Zellulose ausgelöst und in eine Mikrotiter-Platte (engl. *microtiter plate*) übertragen. Die Glasträger waren entweder Epoxy-beschichtete Corning-Träger (Corning Incorporated, USA) oder Aldehyd-beschichtet Genetix- oder Erie-Träger (Genetix GmbH und Erie Scientific GmbH). Auf die Glasträger wurden pro Spot zwei Tropfen à 0.5 nl der Peptidlösungen aufgetragen. Dazu wurden acht piezoelektrische NanoTips (GESIM, Germany) eines non-contact-Nanoplotters (GESIM, Germany) eingesetzt. Die gedruckten Glasträger (die Mikroarrays) wurden für fünf Stunden bei Raumtemperatur liegen gelassen, mit deionisiertem Wasser gewaschen und dann für eine Stunde mit wässriger Hydroxylaminlösung behandelt. Nach weiteren Waschschrten mit Wasser und Ethanol wurden die Mikroarrays unter einem Stickstoffstrom getrocknet. Die resultierenden Mikroarrays können bei 4° C gelagert werden und bleiben mindestens 18 Monate reaktiv. Eine Glasträger-Charge besteht aus identischen Glasträgern, die parallel bedruckt worden sind. Die Zahl der gleichzeitig (parallel) bedruckbaren Glasträger hängt von der Zahl der Plätze im Glasträger-Drucker ab.

### 3.4 Serumantikörper-Bindungsmessung auf Zellulosemembranen (Makroarrays)

Die peptidbeladenen Zellulosemembranen wurden kurz in 70%igem Ethanol geschwenkt und dann dreimal zehn Minuten mit T-TBS (Tris gepufferte Salzlösung, aus dem engl. *Tris-buffered saline* mit 0.05% Tween 20) gewaschen. Anschliessend wurden die Zellulosemembranen 1 h bei Raumtemperatur in Blockierungspuffer (10% Blockierungsreagenz (CRB, Norwich, UK) und 1% Saccharose in T-TBS-Puffer) geblockt. Die geblockten Zellulosemembranen wurden nochmals 10 Minuten mit T-TBS gewaschen und dann über Nacht mit Serum (1:100 in Blockierungspuffer verdünnt) bei 4°C inkubiert. Das Volumen der Inkubationslösung variierte je nach Grösse der Zellulosemembran, die wiederum von der Anzahl Peptidspots abhing. Nach weiteren drei Waschschritten mit T-TBS wurden markierte sekundäre Antikörper (anti-Maus IgG (gegen ganzes Molekül), Peroxidase-gekoppelt, Affinitäts-isoliert, im Schaf produziert, Produktnr. A5906 (Sigma-Aldrich) und anti-Maus IgM ( $\mu$ -Kette spezifisch), in der Ziege produziert, Produktnr. A8786) in einer Konzentration von 1  $\mu$ g/ml Blockierungsreagenz für eineinhalb Stunden bei Raumtemperatur inkubiert. Nach einem weiteren Waschschriff wurde für zwei Minuten Luminol (Sigma-Aldrich) zugegeben, ein Chemolumineszenzsubstrat, das mit der Peroxidase reagiert. Die chemoluminszierenden Signale wurden mit dem LumiImager<sup>TM</sup> (Boehringer Mannheim, Mannheim, Germany) ausgelesen und als *tif*-Dateien gespeichert.

### 3.5 Serumantikörper-Bindungsmessung auf Glasträger (Mikroarrays)

Die Peptidarrays wurden kurz in 100% Ethanol und dann je dreimal zwei Minuten in T-PBS (PBS von engl. *Phosphate buffered saline*, Phosphat-gepufferte Salzlösung mit 0.05% Tween 20 (Polyoxyethylen(20)-sorbitanmonolaurat)) und destilliertem Wasser gewaschen. Bei der Herstellung der Mikroarrays wurden die Glasträger bereits vorbehandelt, um unspezifische Bindungen zu verhindern. Auf einen weiteren Blockierungsschritt konnte daher verzichtet werden. Nach dem Vorwaschen mussten die Glasträger durch Zentrifugation getrocknet werden, damit die Inkubationskammern mit jeweils fünf Unterkammern (GeneFrame<sup>TM</sup>, Abgene, UK) auf der Oberfläche haften blieben. Zur Inkubation wurden die Seren 1:10 mit T-PBS verdünnt; jeweils 45  $\mu$ l der Serumlösung wurden in die Subarrays gegeben und für vier Stunden bei Raumtemperatur inkubiert. Daraufhin wurden die Glasträger mit destil-

liertem Wasser gewaschen. Die sekundären Antikörper anti-mouse IgM-Alexa Fluor 546, Produktnr. A21045 (Invitrogen), und anti-mouse IgG-Alexa Fluor 647, Produktnr. A21235, wurden in einer grossen Kammer, die über alle fünf Arrays reicht (ohne Unterkammern), mit 300  $\mu$ l Inkubationsflüssigkeit in einer 1:100 Verdünnung mit T-PBS für eine Stunde bei Raumtemperatur inkubiert. Nach weiteren Waschschritten, dreimal zwei Minuten mit T-PBS und dreimal zwei Minuten mit destilliertem Wasser, wurden die Glasträger abermals trocken zentrifugiert. Die Fluoreszenzsignale wurden dann mit einem Microarray-Scanner (GenePix 4000B (Molecular Devices GmbH, Ismaning, Germany)) auf den Laser-Wellenlängen 535 und 635 nm ausgelesen und mittels passender Emissionsfilter (rot: 650-690 nm und grün: 550-600 nm) gefiltert. Die Abbildungen wurden als *tif*-Dateien gespeichert.

## 3.6 Datenakquise und Auslesung

Die Signalintensitäten (SI) der einzelnen Spots wurden mit den Programmen *GeneSpotter Version 2.1.2* bis *2.6.0* (Microdiscovery GmbH, Berlin, Germany) oder *GenePix Pro 6 Microarray Image Analysis* ausgelesen und berechnet. Bei beiden Programmen werden Netze aus Kreisen gelegt, welche um jeden der geometrisch angeordneten Spots einen Kreis von mehr oder weniger identischer Grösse mit mehr oder weniger gleichmässigen Abständen in  $x$ - und  $y$ -Richtung legen. Durch manuelles Auflegen der Gitter können die ausgelesenen Bereiche nahezu identisch gewählt werden, was vor allem für die statistische Auswertung angebracht ist. Die Berechnung der SI der einzelnen Spots erfolgt durch Berechnung des Mittelwertes oder Medians der einzelnen Pixel innerhalb des Spot-Kreises (siehe Abbildung 4.14(b)). Bei den Bildern, die von den inkubierten Zellulosemembranen ausgelesen wurden, ist der gewählte Spotradius 3 Pixel, bei den Bildern, die von den inkubierten Glasträgern ausgelesen wurden, ist der Spotradius 6 Pixel (die Auflösung des Auslesens ist bei Glasträgern grösser, weshalb trotz der kleineren Spotgrösse mehr Pixel ausgewertet werden können). Die Radien der aufgelegten Kreise wurden kleiner als die tatsächliche Spotgrösse gewählt, um sicherzustellen, dass auch bei leicht exzentrischen Spots keine Pixel des Hintergrundes mit in die Medianberechnung einbezogen wurden (siehe dazu auch Abbildung 4.14(b)).

### 3.6.1 Signal-Rausch-Verhältnis

Das Signal-Rausch-Verhältnis (SNR, von engl. *signal-to-noise-ratio*) hat je nach Anwendung verschiedene Definitionen. Die Definition des Verhältnisses

der Signale der Peptidspots gegenüber den Hintergrundsignalen der Träger wird oft verwendet. In dieser Arbeit wird jedoch die Definition des Signal-Rausch-Verhältnisses aus der Bildverarbeitung benutzt, welche mit SNR den Ratio des Mittelwertes  $\mu$  gegenüber der Standardabweichung  $\sigma$  bezeichnet.

$$SNR = \mu/\sigma \quad (3.1)$$

wobei der Mittelwert  $\mu = \frac{1}{n} \sum_{i=1}^n x_i$   
und die Standardabweichung (Varianz)  $\sigma = \left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{1}{2}}$

### 3.7 Normierung der Daten

Für alle Datensätze wurden verschiedene Normierungsarten getestet. In der folgenden Zusammenstellung der getesteten und angewandten Normierungen stehen die Vektoren  $v_i$  für die Daten eines Trägers( $i$ ) nach deren Auslesung, wie in Abschnitt 3.6 beschrieben. Dabei steht Träger( $i$ ) sowohl für Glasträger als auch für Zellulosemembran. Die  $v_{i,j}$  stehen entsprechend für einzelne Werte aus  $v_i$ . Normierte Vektoren heissen  $nv_i$  und ihre einzelnen Werte  $nv_{i,j}$  sind die normierten Werte.

- Keine Normierung

$$nv_{i,j} = v_{i,j} \quad (3.2)$$

- Der Median eines jeden Trägers  $i$  ( $MEDIAN(i)$ ) wird von jeder SI ( $v_{i,j}$ ) des Trägers abgezogen

$$nv_{i,j} = v_{i,j} - MEDIAN(v_{i,1}, \dots, v_{i,n}) \quad (3.3)$$

- Der Median eines jeden Trägers  $i$  wird von jeder SI ( $v_{i,j}$ ) des Trägers abgezogen und durch die Standardabweichung  $\sigma_i$  des Trägers dividiert

$$nv_{i,j} = \frac{v_{i,j} - MEDIAN(v_{i,1}, \dots, v_{i,n})}{\sigma_i} \quad (3.4)$$

- Der Vektor  $v_i$  von jedem Träger( $i$ ) wird auf 1 normiert

$$\|nv_i\| = 1 \quad (3.5)$$

wobei  $\|nv_i\| = \frac{1}{|v_i|} \cdot v_i$  und  $|v_i| = \sqrt{v_{i,1}^2 + v_{i,2}^2 + \dots + v_{i,n}^2}$

Die Auswirkungen und Anwendungen der verschiedenen Normierungen sind im Abschnitt 5.1 beschrieben.

## 3.8 Extern akquirierte Daten

Der unabhängige Datensatz der Inkubationen von humanen Seren stammt aus einer Kooperationen: Der Datensatz setzt sich zusammen aus humanen Seren von Pertussis-geimpften gegenüber Placebo-behandelten Menschen (siehe Abschnitt 1.3.2).

### 3.8.1 Pertussis-Seren

Die Daten setzen sich zusammen aus 37 verschiedenen Glasträgern, von denen sieben nur mit sekundärem Antikörper (Cy5-Anti-IgG), zehn mit Placebo-behandelten Seren und 20 mit »Impfstatus DTPA2/5« inkubiert worden sind; alle Datensätze sind mit Anti-IgG (auf Cy5 Wellenlänge) ausgelesen. Die ursprünglichen Glasträger enthielten 3360 SI, nach Abzug der falsch-positiven Signale blieben 2934 SI in je drei Subarrays pro Inkubation, von denen wir den Median gebildet haben (von einer Inkubation existieren nur die Daten von einem Subarray, wodurch die Medianbildung wegfällt).

## 3.9 Datenanalyse und Statistik

### 3.9.1 Hauptkomponentenanalyse (PCA)

Die Hauptkomponenten wurden mit der Funktion *'princomp'* von MATLAB 7.0 berechnet [31].

### 3.9.2 Lineare Diskriminanten Analyse (LDA)

LDA wurde mit den Algorithmen *'lda'* und *'predict.lda'* von R 2.3.1. berechnet [62], [82].

Die Option *'prior'* (steht für vorausgehende, anteilige Wahrscheinlichkeit der einzelnen Gruppen) wurde zur Eliminierung des systematischen Fehlers (engl. *bias*), der infolge der unterschiedlichen Anzahl von Datenpunkten pro Gruppe entstehen würde, immer auf gleichberechtigte Wahrscheinlichkeiten gesetzt, bei zwei Gruppen z. B. auf *prior = 0.5, 0.5*. Multiklassen-Klassifizierung in zwei Ebenen erfolgte mit der Option *dimen = 2* mit der Funktion *predict*.

### 3.9.3 Anpassung an die Normalverteilung

Die Anpassung an die Normalverteilung wurde mit der Funktion *'histfit'* von MATLAB 7.0 berechnet.

### 3.9.4 Korrelation

Die Korrelationskoeffizienten wurden mit der MATLAB-Funktion `'corrcoef'` berechnet.

### 3.9.5 Sensitivität und Spezifität

Sensitivität und Spezifität sind statistische Grössen zur Beurteilung einer binären Klassifizierung oder Vorhersage.

$$Sensitivitaet = \frac{TP}{TP + FN} \quad (3.6)$$

$$Spezifizitaet = \frac{TN}{TN + FP} \quad (3.7)$$

wobei TP = Anzahl richtig Positiver, TN = Anzahl richtig Negativer, FP = Anzahl falsch Positiver und FN = Anzahl falsch Negativer (klassifizierter oder vorhergesagter).

### 3.9.6 Wilcoxon-Vorzeichen-Rang-Tests

Der Wilcoxon-Rangsummentest (engl. *Wilcoxon's matched pair test*) dient dem Vergleich der Mediane zweier Gruppen [27], [21]. Der Test wurde mit der Funktion `'ranksum'` von MATLAB 6.0 durchgeführt. Die berechneten Werte repräsentieren die Wahrscheinlichkeit, dass zwei Proben (in zwei Vektoren) von zwei Verteilungen mit gleichen Median-Werten stammen. In dieser Arbeit wurden die Signalintensitäten von jedem Peptid in jeweils zwei Gruppen miteinander verglichen. Für jedes der Peptide wurde mit dem Wilcoxon-Rangsummentest eine Wahrscheinlichkeit  $p$  berechnet. Die Peptide wurden darauf in der Reihenfolge dieser Wahrscheinlichkeiten  $p_{i...n}$  geordnet. Der kleinste  $p$ -Wert steht dabei für die grösste Wahrscheinlichkeit unterschiedlicher Median-Werte in den zwei verglichenen Gruppen. Oder anders gesagt für die grösste Wahrscheinlichkeit der Unterscheidbarkeit der zwei verglichenen Gruppen mit dem jeweiligen Peptid.

### 3.9.7 Heat maps und Clustering

Heat maps sind grafische Darstellungen von geclusterten Daten als Farben. Die heat maps in dieser Arbeit wurden mit der Funktion `'heatmap.2'` von R 2.6.1 erstellt. `'heatmap.2'` benutzt die Funktion `'cluster'` für das Clustering. Clustering wurde auch mit den MATLAB-Funktionen `'cluster'` und `'link'`

durchgeführt. Ward Clustering beispielsweise ist ein hierarchischer Clusteralgorithmus der immer die zwei Gruppen zusammenclustert bei deren Fusion der minimalste Informationsverlust bezüglich des ganzen Datensatzes entsteht.

### 3.9.8 Multidimensionale Skalierung

Die multidimensionale Skalierung wurde mit der Funktion *'cmdscale'* von R 2.6.1 berechnet.

### 3.9.9 Potential Support Vector Machine (P-SVM)

Die Potential Support Vector Machine (P-SVM) [26], [34] ist eine neuere, überwachte Maschinen-Lernmethode, die mit denselben Kernels wie gewöhnliche Support Vektor Maschinen benutzt werden kann. Des Weiteren kann P-SVM direkt auf dyadische (zweiteilige) Datensätze angewandt werden, was z. B. in unserem Fall rechteckige Matrizen sind, mit den Signalintensitäten zweier Gruppen gegenüber aller gemeinsam vorkommender Peptiden. Die Einträge der Matrizen beschreiben die Beziehung der klassenzugehörigen Datensamples (Seren) zu jenen der Proben (Peptide), die als Deskriptoren (Merkmale) der Samples dienen. Das Schema, das die P-SVM regelt, selektiert kleine informative Probensets, die zur Merkmalsselektion benutzt werden können. Die maximale Anzahl auszuwählender Merkmale für die P-SVM Berechnung kann frei gewählt werden. Um eine minimale Anzahl von Peptiden mit einer maximalen Klassifizierungs- bzw. Vorhersage-Fähigkeit zu erhalten, macht es Sinn, für alle Merkmalsanzahlen von Interesse die P-SVM separat zu berechnen (hat die praktische Erfahrung gezeigt).

Bei der Kreuzvalidierung werden immer  $n-1$  Datenpunkte zur Merkmalsauswahl genommen und der  $n$ -te Datenpunkt mittels dieser Merkmale dann vorhergesagt; wobei  $n$  der Anzahl der Samples beider Gruppen entspricht. Dieses Vorgehen wird  $n$  Mal durchgeführt, bis jeder Datenpunkt einmal vorhergesagt wurde. Dabei werden für unterschiedliche  $n-1$ -Vorhersage-Datensets leicht unterschiedliche Merkmalsgruppen gewählt. Nach allen  $n$  Läufen wird die Häufigkeit jedes Merkmals berechnet und die repräsentativsten werden ermittelt. Gemäss der Einteilung der Merkmalsselektions-Methoden in Abschnitt 1.7.2 gehört die P-SVM am ehesten zu den »Wrapper« Methoden.

### 3.9.10 Bestimmung der Signifikanzen der Resultate

Um die Signifikanz der Präzision der Klassifizierung oder Kreuzvalidierungs-Vorhersage mit der gewählten Anzahl von Merkmalen zu testen, werden die Labels (Gruppenkennzeichnungen) der beiden Gruppen zufällig vertauscht. Der dadurch entstandene, vermischte Datensatz wird mit P-SVM auf ebenso gute oder bessere Klassifizierung bzw. Vorhersagbarkeit getestet. Die Anzahl der ausgetauschten Gruppenkennzeichnungen ist immer mindestens halb so gross wie die Hälfte der Anzahl der Daten in der kleineren Gruppe. Dieses Austausch- und Neuberechnungsverfahren wird so oft wiederholt, bis die gewünschte Genauigkeit der Signifikanz erreicht wird (hundert Wiederholungen mit nur schlechteren Klassifizierungs-, oder Vorhersage-Resultaten ergeben dann einen p-Wert der Signifikanz von  $p < 0.01$ ).



# Kapitel 4

## Optimierung und Evaluation der Technologien

Das Kapitel beschreibt die Resultate der Experimente und Vorversuche zur Evaluation der Grundlagen der Serumantikörper-Reaktivitäts-Profilen-Messungen auf Zellulosemembranen und Glasträgern.

Mittels Serumantikörper-Reaktivitäts-Profilen könnte eine ganze Reihe interessanter Fragestellungen untersucht werden. Die beschränkte Verfügbarkeit von Rohmaterialien reduzierte im praktischen Alltag allerdings die Zahl der Möglichkeiten. So ist beispielsweise der beschränkte Zugang zu Seren-samples in genügend grosser Zahl und guter Charakterisierung ein limitierender Faktor. Die Herkunft der Seren ist in Abschnitt 3.1 beschrieben. Die Auswahl der synthetischen 14- und 15mer Peptidsequenzen wird in Abschnitt 3.2 beschrieben. Das Protokoll ihrer Synthese findet sich in Abschnitt 3.3, wobei einzelne Problempunkte in den Abschnitten 4.2.3 – 4.2.6 besprochen werden.

Die Signalintensitäten der Bindungsmessungen der Serumantikörper gegenüber Peptidbibliotheken waren bei beiden Trägern grossen Abweichungen und Varianzen unterworfen. Die Varianzen entstanden aufgrund technologischer Unbeständigkeiten. Die grössten technologischen Schwankungen entstanden sowohl bei den Zellulosemembranen wie auch bei den Glasträgern zwischen unterschiedlichen Chargen. Als Chargen bezeichnet man Träger, die in parallelen Prozessen hergestellt werden. Die genaue Definitionen der Chargen findet sich in den Abschnitten 3.3.2 und 3.3.3, siehe auch Abbildung 4.5. Schwankungen sind aber auch zwischen einzelnen Versuchen, Tagen, Experimentatoren und wahrscheinlich einigen anderen Faktoren vorhanden.

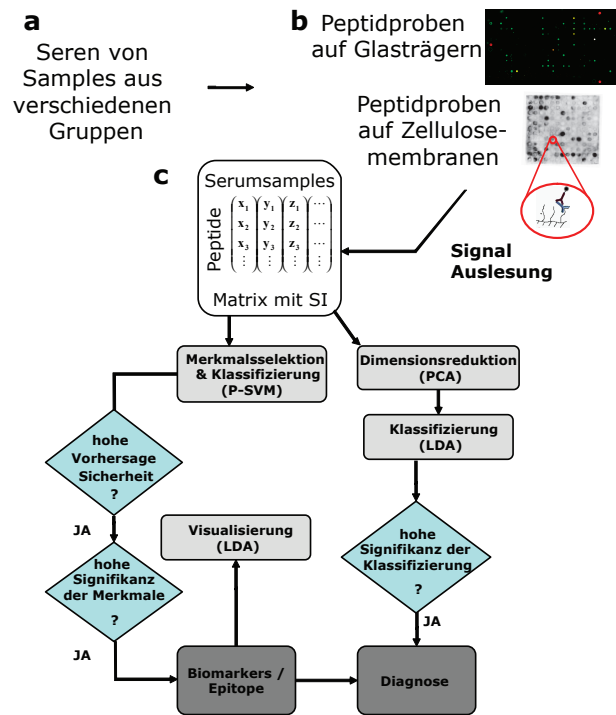
Die Untersuchungen der Serumantikörperbindungen erfolgte im ersten Teil der Arbeit auf Zellulosemembranen und in einer zweiten Phase auf Glasträgern. Die Trägertechnologie wurde geändert, da es ab 2006 technologisch möglich war, Inkubationen auf Glasträgern mit sinnvollem Signal-Rausch-Verhältnis (SNR, siehe Abschnitt 3.6.1 für Definition) hervorzubringen (siehe Abbildung 4.8). Ein direkter Vergleich der Bindungen auf Zellulosemembran und Glasträgern erscheint nicht sinnvoll, weil sich die Längen der verwendeten Peptide unterscheiden (15mere auf Zellulosemembranen und 14mere auf Glasträgern). Ein Vergleich beider Technologien folgt in Abschnitt 8.5.

## 4.1 Übersicht der verwendeten Technologien und Prozesse

Die Arbeit besteht aus einem experimentellen und einem analytischen Teil. Abbildung 4.1 zeigt eine Übersicht der Prozesse, die benutzt wurden, um die Antikörper-Reaktivitäts-Profile verschiedener Gruppen von Serensamples zu vergleichen. Untersucht wurden jeweils zwei Serengruppen von Mäusen (Kapitel 5 und 6) oder Menschen (Kapitel 7). Die Serenproben wurden auf unterschiedlichen Trägern inkubiert (Abbildung 4.1(b)). Kapitel 5 behandelt die Inkubation der Serenproben auf Zellulosemembranen, die Kapitel 6 und 7 analysieren die Inkubationen der Seren auf Glasträgern. Die Zellulosemembran in Abbildung 4.1(b) enthält  $13 \times 13 = 169$  Peptide, während der Subarray des Glasträgers 300 unterschiedliche Peptide (von denen 45 nicht beachtet wurden) und zusätzlich Kontroll-Spots enthält. In den Ecken der Glasträger-Subarrays befinden sich weitere Markierungen: TAMRA als Fluoreszenzkontrolle sowie murine IgM- und IgG-Antikörper als Kontrolle der sekundären Antikörperbindungen. Die Serumantikörperbindungen werden mittels markierter sekundärer Antikörper sichtbar gemacht. Auf Zellulosemembranen sind die sekundären Antikörper mit POD-Enzymen, auf den Glasträgern mit Fluorophoren markiert. Für die murinen Sekundärantikörper gilt: Auf den Zellulosemembranen wurden nur IgM-Antikörperbindungen in die Auswertungen einbezogen. Auf den Glasträgern emittieren die Fluorophore der anti-IgM-Antikörper im grünen Bereich (anti-IgM-Alexa Fluor 546), und die der anti-IgG-Antikörper im roten Bereich (anti-IgG-Alexa Fluor 647).

Der Weg der statistischen Analyse wird in Abbildung 4.1(c) dargestellt: Die ausgelesenen IgM-Signale der Zellulosemembranen wurden normiert, die der Glasträger nicht (mit Ausnahme von extern akquirierten Daten, siehe Kapitel 7). Die Hauptkomponentenanalyse (PCA) wird zur Dimensionsreduk-

tion benutzt, indem sie die höchsten Varianzen extrahiert und dadurch einen Eindruck der Sampleverteilung im zweidimensionalen Raum verschafft (erste Hauptkomponente vs. zweite Hauptkomponente aufgetragen). Gut klassifizierbare Datensets (bzw. Datensets, die aufgrund ihrer höchsten Varianzen klassifizierbar sind) können bereits im PCA-Plot eine Auftrennung aufweisen (wie z. B. die Pertussis-geimpfte gegenüber der mit Placebo behandelten Gruppe in Kapitel 7). Datensätze, die nicht aufgrund der höchsten Varianz auftrennbar sind, erfordern »voreingenommene« Methoden zur Klassifizierung (wie die Lineare Diskriminanten Analyse (LDA)), wozu teilweise vorgängige Merkmalsselektion nötig ist (z. B. mittels des Algorithmus der Potential Support Vector Machine (P-SVM)).



**Abbildung 4.1:** Übersicht: Inkubationen der Serumantikörper-Inkubationen und der Analysen der resultierenden Bindungssignalen auf Peptidbibliotheken. Experiment und Analyse-Schritte zur Klassifizierung verschiedener Gruppen (z. B. von gesunden und kranken) anhand von Serumantikörper-Reaktivitäts-Profilen. (a) Verschiedene Gruppen von Seren werden ausgewählt. (b) Die Inkubationen auf verschiedenen Trägern (Zellulosemembranen oder Glaträger) und die Bindungsdetektion. (c) Die Auswertung und Analyse der SI-Matrizen der untersuchten Serengruppen verläuft nach den aufgezeichneten Schemata.

## 4.2 Grundlagen der Technologien

Um eine Technologie sinnvoll anzuwenden und zu interpretieren ist es notwendig, die idealen Anwendungsparameter und die Zuverlässigkeit dieser Technologie in Erfahrung zu bringen. In unserem Fall beinhaltete die Evaluation der Anwendungsparameter unter anderem das Testen der Konzentrationen der Reagenzien, der Protokollzeiten, der geeigneten Markierungen und Detektion der sekundären Antikörper sowie das Testen verschiedener Herstellungsverfahren und Trägermaterialien. Die Zuverlässigkeit der Technologie steht dabei als Mass für die Reproduzierbarkeit der gemessenen Ergebnisse, welche wiederum von Umweltfaktoren wie Temperatur, Luftfeuchtigkeit, Experimentatoren-Wechsel und anderen Faktoren abhängen können. Der Evaluation der Technologien wurde bei der vorliegenden Arbeit grosses Gewicht beigemessen.

### 4.2.1 Detektion von Serumantikörperbindungen

Zur Untersuchung der Inkubation von Serumantikörpern auf Zellulosemembran mussten erst grundlegende Fragen geklärt werden, wie:

- Können Serumantikörperbindungen auf systematischen (zufälligen) Peptidbibliotheken nachgewiesen werden?
- Was sind die optimalen Inkubationsverdünnungen der Seren und sekundären Antikörper?
- Gibt es geeignetere Sekundär-Antikörper-Markierungen als POD, welches für den Nachweis der Bindung monoklonaler Antikörper auf Zellulosemembranen verwendet wird?

Die Vorversuche lieferten folgende Ergebnisse:

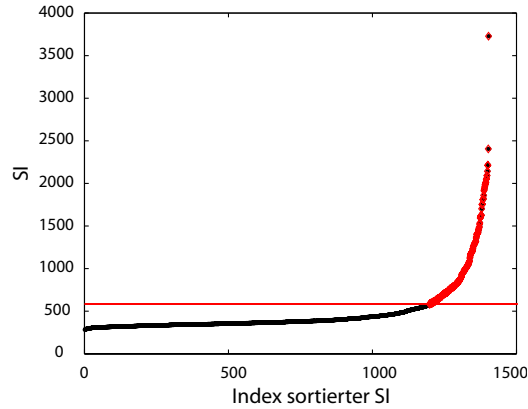
**Nachweis von Serumantikörperbindungen** Ein Viertel der ersten Membran (mit 1403 Peptiden) wurde mit gemischten (engl. *pooled*) Seren von drei BALB/c-Mäusen inkubiert. Die Bindungen der Serumantikörper wurden mit dem POD-markierten anti-Maus-IgG-Sekundärantikörper A5906 nachgewiesen. Gemäss Abbildung 4.2 wiesen 14.7% aller Peptide IgG-Ak-SI über dem abgeschätzten Hintergrundsignal (das hier als minimale SI plus Standardabweichung aller SI definiert ist) auf. Das Resultat zeigte, dass durch Inkubation von Seren auf systematischen Peptidbibliotheken (Abschnitt 1.5.2) auf Zellulosemembran nachweisbare SI entstehen. Bei der Analyse der Reaktivitäts-Profile auf Zellulosemembranen wurden dann ausschliesslich IgM-Ak-SI untersucht (Abschnitt 5), während auf Glasträgern die IgM- und IgG-Ak-SI parallel

ausgelesen werden konnten. In den SI-Histogrammen von anderen Zellulosemembranen (Abbildungen 4.6) und von Glasträgern (Abbildungen 4.10) liegt ebenfalls ein grosser Anteil der SI im 'Hintergrundsignal'-Bereich. Nur ein kleiner Anteil der Peptidspot-SI-Messungen hebt sich davon ab.

**Geeignete Konzentrationen der Seren und Antikörper** Um die Signale der Bindung der Serumantikörper auf der systematischen Peptidbibliothek auf Zellulosemembran gut sicht- und auswertbar zu machen, erwies sich eine Verdünnung der Seren 1 : 100 in Blockingbuffer am geeignetsten (wie in Abschnitt 3.4 beschrieben). Die sekundären Antikörper wurden in einer Konzentration von 1  $\mu\text{g}/\text{ml}$  in Blockingbuffer nach Standardprotokoll für monoklonale Antikörperdetektion zugegeben. Die Anwendung von kleineren Konzentrationen resultierte in schwächeren Signalen (d.h. schlechteres Signal-Rausch-Verhältnis), während grössere Konzentrationen Bilder mit »Fleckenbildung« hervorriefen, die wohl auf ineinander verklumpte Antikörper zurückzuführen waren (Daten nicht gezeigt). Für Seruminkubationen auf Glasträger erwies sich die kleinere Verdünnung der Seren 1 : 10 in Blockingbuffer als geeigneter. Die Seruminkubationen auf Glasträger erfordern grössere Konzentrationen, weil die Peptiddichte aufgrund der Immobilisierungschemie auf den Glasträgeroberflächen um ein Vielfaches kleiner ist, was auch der Vergleich der Peptidkonzentrationen/ $\text{mm}^2$  in Tabelle 1.1 zeigt. Die Antikörperkonzentrationen der Seren wurden weder bestimmt noch angepasst.

**Detektion mittels fluoreszierendem Farbstoff** FITC-gekoppelte Anti-IgG-Antikörper (Produkt Nummer F0257, Sigma Aldrich) in der Originalkonzentration von 1.1  $\text{mg}/\text{ml}$  wurden in Konzentrationen von 1, 2, 5 und 10  $\mu\text{g}/\text{ml}$  Blockierungsreagenz getestet und auch direkt in kleinen Tropfen auf Membranstücke aufgetragen (ohne waschen) und auf Sichtbarkeit getestet. Im Fluoreszenzmodus des LumiImager<sup>TM</sup>-Gerätes waren keine Signale detektierbar, die höher fluoresziert haben als die Zellulosemembran selbst. Auch andere getestete Fluorophore ergaben keine Fluoreszenzsignale, die eine höhere SI als die Eigenfluoreszenz der Zellulose zeigten. Die Bindungsmessung mit fluoreszierenden Farbstoffen machte dafür auf den Glasträgerarrays sehr viel Sinn. Glasträger haben keine Eigenfluoreszenz. Die Messung der Fluoreszenz verlangt allerdings sehr homogene Oberflächen und eine sorgfältige Inkubationstechnik: kleinste Verunreinigungen (auch in den Inkubationsreagenzien) und Kratzer waren sehr gut sichtbar, siehe z. B. Abbildungen

4.7(b) und 4.8(a) und 'überstrahlten' die Ak-Bindungssignale.



**Abbildung 4.2:** Sortierte Signalintensitäten eines Serenpools, der auf einem Stück der ersten Membran mit 1403 Peptidspots inkubiert wurde. Die schwarzen Punkte sind die Signalintensitäten einzelner Peptidspots, die rote Linie ist ein frei definierter »Hintergrund-Grenzwert« und die roten Rauten markieren darüberliegende Peptidspot-SI.

#### 4.2.2 Genauigkeit der Bindungsmessungen: Binder und Nicht-Binder

Für den monoklonalen Antikörper CB4-1 und seine bekannten Binder wurde die Genauigkeit der Bindungsmessung an voreingenommenen Bibliotheken auf Zellulosemembranen und Glasträgern in den Arbeiten von Weiser u. a. [85] und Tapia u. a. [77] analysiert. Beide Arbeiten kommen zum Schluss, dass die Bindung von monoklonalen Antikörpern auf Peptidarrays mit unterschiedlichen bekannten CB4-1 Bindern nur zur Unterscheidung von Bindern und Nichtbindern dient, und dass die SI nur schlecht mit den Dissoziationskonstanten  $K_d$  korrelieren. Auf den Zellulosemembranen liegt der Grenzwert zwischen Binder- und Nichtbinder-Klasse bei  $10^{-6}$  [85], was für die Stärke der Antikörperbindung eine physiologisch relevante Grösse ist. Auf den Glasträgern liegt dieser Grenzwert gar bei  $10^{-7}$  [77]. Die untersuchten Kovarianzen waren bei beiden Technologien in der gleichen Grössenordnung. Auf Glasträgern war die Reproduzierbarkeit der SI der monoklonalen Ak deutlich grösser (innerhalb von Chargen) und direkt vergleichbar, während die SI der Zellulosemembranen nur nach vorhergehender Normalisierung vergleichbar

waren. Ein Ergebnis, das in der vorliegenden Arbeit für Serumantikörper-SI durch Vergleich der Korrelationskoeffizienten und mittleren SNR bestätigt wird (siehe Abschnitte 5.1 und 6.1).

### 4.2.3 Synthese von Cystein-enthaltenden Peptiden

Cystein ist berüchtigt, bei der chemischen Synthese von Peptiden Probleme zu bereiten und dadurch bei Interaktionsmessungen die Bindung der Antikörper zu unterbinden. Zur Untersuchung, ob Peptide, die Cystein (Cys) enthalten, schlechter synthetisiert werden können, hatten wir auf den Membranen aus Charge vier alle Cys ausgewechselt. Diese Membranen enthalten folglich 174 Peptide mehr als die der vorherigen Chargen. 126 Peptide dieser Differenz stammen aus dem Vergleich der Aminosäuren-Sequenzen mit und ohne Cystein (Cys). Alle Peptide, die Cys enthielten wurden zusätzlich durch identische Peptide mit Serin (Ser) anstelle von Cys ersetzt, was 145 AS-Austauschen in 126 Peptiden entsprach. Dadurch wurden alle Peptide, die Cys enthalten doppelt synthetisiert. Die verglichenen Peptide wurden in zwei Gruppen aufgeteilt: Gruppe ( $C \rightarrow S$ ) sind die Peptide mit Ser statt Cys und Gruppe ( $C$ ) diejenigen mit Cys. Der Korrelationskoeffizient der geordneten Gruppen ( $C \rightarrow S$ ) und ( $C$ ) betrug  $S = 0.73$ . Zum Vergleich wurden auch 30 Repetitionen von je vier verschiedenen Peptiden aufgetragen, deren mittlerer Korrelationskoeffizient eher schlechter war (d.h.  $S < 0.73$ ). Daraus folgerten wir, dass es zwischen den Peptiden, die Ser oder Cys enthalten, keinen erheblichen Unterschied in der Erkennung gibt. In allen folgenden Synthesen der Zellulosemembran-Makroarrays und auch in den Peptidsynthesen für die Glasträger-Mikroarrays wurden die Cys enthaltenden Peptidsequenzen nicht weiter beachtet.

### 4.2.4 Regeneration der Zellulosemembranen

Die Regeneration bereits inkubierter Membranstücke (mit Stücken der Membran aus Charge 1, siehe Tabelle 4.1) ergab keine befriedigenden Resultate. Die Behandlung nach Standardprotokollen der Membranregeneration [61] erzielte nach Inkubation mit Luminol keine sichtbaren Signale mehr. Neu inkubierte Seren brachten aber deutlich weniger Signale auf der regenerierten Membran hervor. Aufgrund des Verdachtes, dass die Regeneration nach Standardprotokoll auch die Peptide auf der Membran angreifen könnte, wurde die Membran in einem weiteren Regenerationsversuch nur mit  $H_2O$  erhitzt (10 min bei  $50^\circ C$ , 15 min bei  $70^\circ C$  und nochmals 20 min bei  $70^\circ C$ , dreimal TBS und dazwischen einmal mit  $H_2O$  gewaschen), was aber zu gleichen Ergebnis-

sen, wie die Regeneration mit Standardprotokoll führte.

Die Abwesenheit von Signalen spricht aber nicht eindeutig für eine Regeneration der Membran, sondern kann auch bedeuten, dass lediglich die Enzymfunktion der Peroxidase zerstört worden ist, oder nur der zweite Antikörper zerstört und/oder eliminiert wurde. Aus diesen Gründen wurde von der Regeneration der Membranen abgesehen.

#### **4.2.5 CAPE-Membranen und Beta-Ala-Membranen im Vergleich**

Die SPOT-Synthese erlaubt die Verwendung verschiedener Linker zur Funktionalisierung der Zellulosemembranen (Abschnitt 3.3.1). In dieser Arbeit wurden nur die Synthesen mittels  $\beta$ -Ala- und CAPE-Linker untersucht (abgesehen von den JPT-eigenen, nicht bekanntgegebenen Linkern zur Synthese der Peptide für die Glaträger). Der Ansatz des Vergleichs der beiden Technologien wurde durch die Herstellung der dritten Charge der Zellulosemembranen mit der CAPE-Methode versucht. Die Korrelationen der Signale von Seruminkubationen aus dem gleichen Mausstamm zwischen Inkubationen auf der zweiten und der dritten Charge waren deutlich tiefer als die Korrelationen der SI innerhalb der Stämme aus Tabelle 4.2 (Resultat nicht gezeigt). Die Resultate der CAPE-Membranen wurden verworfen. Bei der Synthese der folgenden Chargen von Zellulosemembranen wurde nur noch die  $\beta$ -Ala-Technologie verwendet (Tabelle 4.1).

#### **4.2.6 Grössere Zahl identischer Membranen pro Charge – Reduktion der Anzahl der Peptide**

Die kleineren Korrelationen zwischen den unterschiedlichen Chargen (Abschnitt 4.3) veranlasste uns zur Entscheidung, die Zellulosemembran-Chargen mit weniger Peptiden, dafür in grösserer Zahl von identischen Membranen pro Charge herzustellen. Dazu musste die Anzahl der Peptide von rund 440 auf rund 150 reduziert werden (siehe Abschnitt 3.3.2).

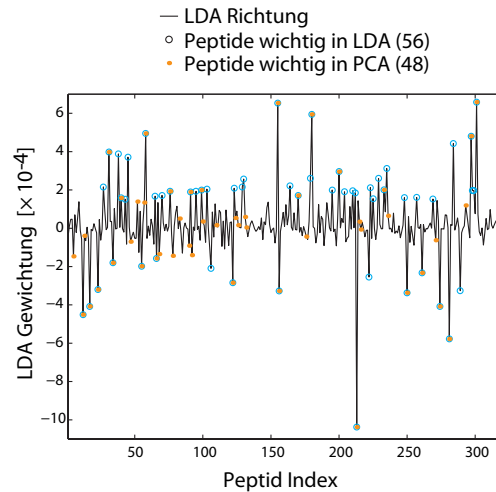
#### **Selektion entscheidender Peptide mittels LDA-Richtung**

Die Entscheidung, pro Charge eine grössere Zahl von kleineren Membranen herzustellen, verlangte eine Reduktion der Anzahl der Peptide pro Membran. Eine Auswahl von Peptiden erfolgte mittels der LDA-Gewichtung der Peptide, die am besten zwischen den Gruppen der BALB/c- und C57BL/6-Seren der zehn inkubierten Membranen aus Charge zwei und vier unterscheiden



können. Die 10% der Peptide, die am meisten zur LDA-Separation beitragen (d.h. 56), wurden mit 100 zufällig gewählten Peptiden ergänzt, um Möglichkeiten für andere Bindungsunterschiede von bis dahin unbeachteten Differenzen zu bieten. Abbildung 4.3 zeigt die LDA-Gewichtungen der Peptide und markiert die 56 wichtigsten Peptide.

Die gewählten 156 Peptide wurden für die folgenden synthetisierten Membran-Chargen beibehalten. Alle Klassifizierungen der Zellulosemembrandaten beziehen sich auf die 156 Peptide, die auf allen ausgewerteten Membranen vorkommen und im Anhang aufgelistet sind.



**Abbildung 4.3:** Mittels LDA-Richtung selektierte Peptide. Die 56 Peptide mit der maximalen Gewichtung in der LDA-Klassifizierung von BALB/c- und C57BL/6-Signalen. Die Überschneidung mit den Peptiden, die zur maximalen Varianz in PCA beitragen, ist offensichtlich.

Die Zahl der Peptide auf den Glaträgern wurde im Verlauf der Zeit ebenfalls reduziert, weil das Gitter zum Auslesen der Peptidspots auf kleineren Arrays viel einfacher und genauer aufgelegt werden kann. Die Reduktion der Peptidsequenzen der Glaträger geschah aufgrund der Diversität der geplanten weiteren Experimente mittels Zufallsgenerator.

#### 4.2.7 Verzerrung der experimentellen Ergebnisse durch Versuchsanordnung

Die Samples der verschiedenen murinen Gruppen wurden nicht getrennt behandelt. Auf Durchmischung der Gruppen wurde innerhalb der Träger-Chargen, der Versuchstage und der Subarrays der Glasträger geachtet. Denn aufgrund der eher geringen Reproduzierbarkeiten der Technologien ist es oft einfacher Merkmale zu finden, welche die Experimente von einzelnen Tagen unterscheiden, als Merkmale zur Aufdeckung der biologischen Differenzen.

### 4.3 Reproduzierbarkeit der Serumantikörperbindungs-Signale an Peptidarrays auf Zellulosemembranen

In unserer Arbeit haben wir wiederholte Inkubationen des gleichen Serums auf unterschiedlichen Zellulosemembranen, aufgrund der geringen Zahl manuell hergestellter Membranen, nur zwischen manuell und vollautomatisch hergestellten Membranen (d.h. auf unterschiedlichen Chargen) und zwischen vollautomatisch hergestellten Membranen aus der gleichen Charge durchgeführt. (Zur unterschiedlichen Herstellung der Membranen siehe Abschnitt 3.3.2).

Tabelle 4.1 zeigt eine Übersicht der hergestellten Chargen und die Zahlen der Membranen pro Charge. Alle Membranen wurden in der *Arbeitsgruppe Molekulare Bibliotheken am Institut für medizinische Immunologie des Universitätskrankenhauses Charité Berlin* im Rahmen einer Kooperation hergestellt.

Die Inkubationswiederholungen haben überraschend niedrige Korrelationskoeffizienten: Für die 156 gemeinsamen Peptide von drei Inkubationswiederholungen eines C57BL/6-Serums auf Membranen der vierten, fünften und siebten Charge, ist der mittlere Korrelationskoeffizient  $S = 0.43$  und der mittlere SNR  $= 2.23$ . Der Korrelationskoeffizient von drei Inkubationswiederholungen eines anderen C57BL/6-Serums auf drei Membranen der siebten Charge beträgt immerhin  $S = 0.67$ , der mittlere SNR  $= 7.12$  (Definition SNR siehe Abschnitt 3.6.1). Die Mittelwerte und Standardabweichungen der beiden dreifachen Inkubationswiederholungen sind in Abbildung 4.4 gezeigt. Die Reproduzierbarkeit der Bindungsmessungen ist also eher gering, liesse

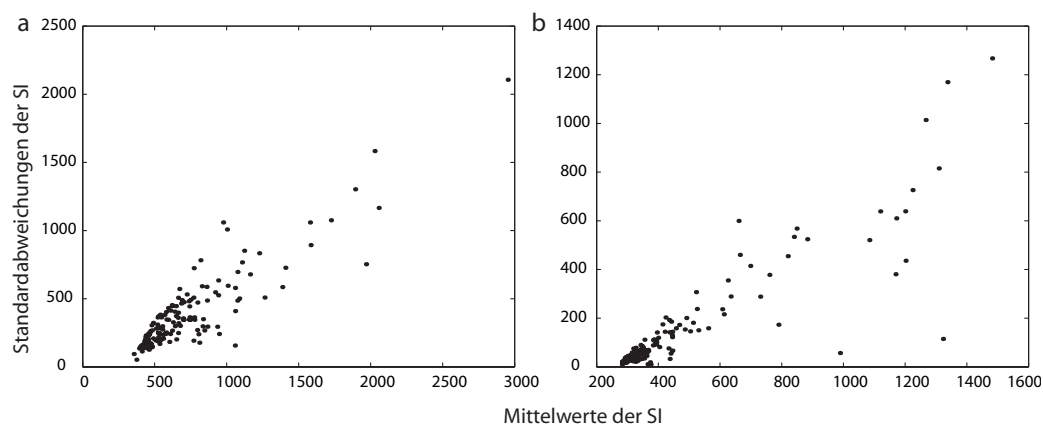
Charge	Linker	Herstellung	Anzahl Peptide	Anzahl Rep.	in Res.	Datum
1	$\beta$ -Ala	manuell	5520	1	Nein	April 2003
2	$\beta$ -Ala	manuell	508	6	Ja	Sept. 2003
3	CAPE	manuell	682	2	Nein	Jan. 2004
4	$\beta$ -Ala	manuell	682	4	Ja	Mai 2004
5	$\beta$ -Ala	manuell/auto	156	$2 \times 12$	Ja	Aug. 2004
6	$\beta$ -Ala	auto	169	20	Ja	Feb 2005
7	$\beta$ -Ala	auto	169	20	Ja	März 2005
8	$\beta$ -Ala	auto	169	11/20	Ja	Juni 2005
9	$\beta$ -Ala	auto	169	13/20	Ja	Dez 2005

**Tabelle 4.1:** Die Zellulosemembranen, die im Rahmen dieser Arbeit hergestellt und inkubiert wurden. »Linker« bezeichnet das Verbindungsstück, mit dem die Peptide an die Aminofunktionen der Zellulosemembran derivatisiert wurden:  $\beta$ -Ala steht für *mit  $\beta$ -Alaninbaustein Fmoc- $\beta$ -Ala-OPfp derivatisiert* und CAPE steht für *Zelluloseaminopropylether-derivatisiert*.

sich jedoch bei besserer Verfügbarkeit manuell hergestellter Träger erhöhen.

Repetitionen auf manuell hergestellten Membranen ergäben deutlich besser korrelierende Ergebnisse, wie der Vergleich zu den Korrelationen unterschiedlicher Individuen aus zwei Stämmen in Tabelle 4.2 zeigt: Die SI von Inkubationen von Seren aus dem gleichen Stamm korrelieren sehr hoch und die SI von Sereninkubationen aus unterschiedlichen Stämmen deutlich tiefer. Die hohen Korrelationen der SI zwischen Seren des gleichen Mäusestammes legen nahe, dass die Korrelationskoeffizienten zwischen wiederholten Inkubationen des gleichen Serums auf manuell hergestellten Zellulosemembranen deutlich höher wären. Leider war es jedoch nicht möglich, eine genügend grosse Zahl manuell hergestellter Membranen oder -chargen für Wiederholungsversuche zu erhalten; jedenfalls nicht in sinnvoller Zeit. Deshalb haben wir auf vollautomatisch hergestellte Membran-Chargen umgestellt.

Trotz der hohen Korrelationen zwischen den SI von Sereninkubationen aus dem gleichen Stamm sehen die inkubierten Membranen sehr unterschiedlich aus. Abbildung 4.5 illustriert die resultierenden Signale der Membranen einer Charge (Charge zwei). Die Membranen sind mit Seren aus nur zwei unterschiedlichen Mäusestamm-Gruppen (gesunde BALB/c und C57BL/6) inkubiert worden, welche sich aufgrund der visuell erkennbaren Muster nicht mehr voneinander unterscheiden, als die Unterschiede innerhalb der Stämme



**Abbildung 4.4:** Mittelwerte und Standardabweichungen der Signalintensitäten von 156 Peptiden von jeweils drei Inkubations-Repetitionen: (a)  $\mu$  und  $\sigma$  von drei Inkubationen eines C57BL/6-Serums auf Membranen der vierten, fünften und siebten Charge. (b)  $\mu$  und  $\sigma$  von drei Inkubationen eines C57BL/6-Serums auf drei Membranen der siebten Charge.

unterschiedlich zu sein scheinen. Als Ergänzung zu diesen Analysen zeigt Abbildung 4.6 die Verteilung der IgM-Ak-Signale auf einer Zellulosemembran der zweiten Charge.

## 4.4 Reproduzierbarkeit der Serumantikörperbindungs-Signale an Peptidarrays auf Glasträgern

Die Glasträgertechnologie entwickelte sich erst im Verlauf dieser Arbeit zu einer experimentellen Plattform mit guter Reproduzierbarkeit. Wir haben sehr viele Versuche mit sehr variierenden Ergebnissen vorgenommen, wobei aufgrund der grossen Empfindlichkeit der Glasoberfläche nicht immer klar war, ob Ergebnisse mit grossem Rauschen von Fehlern in der Oberflächenbeschichtung, in der Peptidauftragung oder im Inkubationsprotokoll zu optimieren waren.

	B 1	B 2	B 3	B 4	B 5	C 1	C 2	C 3	C 4	C 5
B 1	1.00	0.80	0.71	0.62	0.55	0.55	0.54	0.21	0.20	0.42
B 2	0.80	1.00	0.80	0.72	0.58	0.41	0.39	0.20	0.21	0.45
B 3	0.71	0.80	1.00	0.88	0.74	0.36	0.37	0.27	0.18	0.58
B 4	0.62	0.72	0.88	1.00	0.74	0.31	0.34	0.32	0.24	0.64
B 5	0.55	0.58	0.74	0.74	1.00	0.38	0.37	0.39	0.32	0.62
C 1	0.55	0.41	0.36	0.31	0.38	1.00	0.91	0.63	0.65	0.55
C 2	0.54	0.39	0.37	0.34	0.37	0.91	1.00	0.61	0.61	0.56
C 3	0.21	0.20	0.27	0.32	0.39	0.63	0.61	1.00	0.83	0.62
C 4	0.20	0.21	0.18	0.24	0.32	0.65	0.61	0.83	1.00	0.54
C 5	0.42	0.45	0.58	0.64	0.62	0.55	0.56	0.62	0.54	1.00

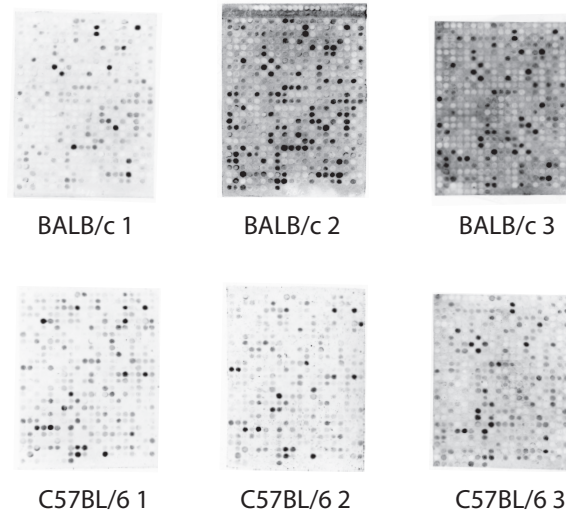
**Tabelle 4.2:** Die Korrelationskoeffizienten der normierten SI der 440 gemeinsamen Peptide der ersten zehn Membranen. B steht für BALB/c und C für C57BL/6, individuelle Mäuse wurden innerhalb der Gruppen durchnummeriert.

#### 4.4.1 Unbeständigkeit der Signale von Glasträgern

Hier eine kurze Zusammenfassung der Versuche und Überlegungen, die wir in der Hoffnung auf mehr Verständnis der Technologie unternommen hatten.

Die ersten Seruminkubationen wurden auf Peptiden auf Glasträgern vorgenommen, die uns JPT GmbH (damals noch Jerini AG) aus den Chargen 163, 176, 340, 363, 364 und 369 zu Testzwecken zur Verfügung stellten. Die Glasträger waren mit jeweils 768 Peptiden in drei Wiederholungs-Blöcken auf Genetix-Trägern (siehe Abschnitt 3.3.3) bedruckt. Die Inkubationen erfolgten zwischen April 2004 und Februar 2005 mit verschiedenen BALB/c- und C57BL/6-Seren in unterschiedlichen Konzentrationen. Die resultierenden Signale wiesen ein sehr hohes Rauschen auf (d.h. Schlieren und Schneesturm- oder willkürliche Sternenhimmel-Muster, die nichts mit den regelmässig angeordneten Peptidspots gemein hatten), waren sehr unterschiedlich und meistens nicht auswertbar.

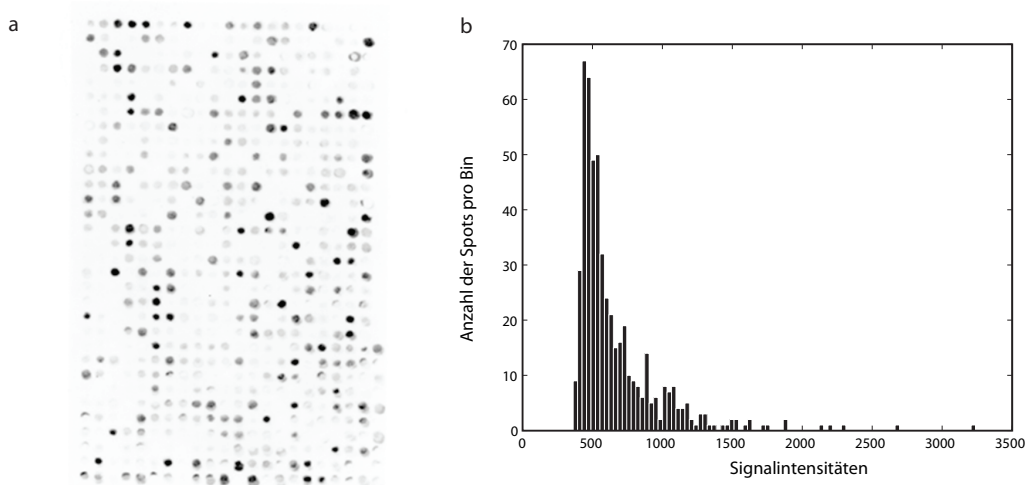
Mit dem Ziel, mehr Repetitionen pro Peptid zu erhalten, wurde die Charge 380 mit 30 mal 40 Peptiden bedruckt. In Abbildung 4.7 ist ersichtlich, dass die 1200 Peptide in einem Block angeordnet waren. Die Anordnung der vielen Repetitionen in einem Block geschah in der Hoffnung, technologische Ursachen der grossen Unterschiede zu finden. Die Signale der Inkubationen auf den Trägern der 380er-Charge waren ebenfalls von sehr unterschiedlicher Qualität. Abbildung 4.7(a) zeigt eines der besten Beispiele dieser Charge,



**Abbildung 4.5:** Die Bilder der inkubierten Zellulosemembranen aus der zweiten Charge. Die abgebildeten Membranen sind identische Subarrays einer grossen Membran. Sie wurden alle in parallelen Arbeitsschritten hergestellt und haben an den gleichen Positionen die gleichen Peptide. Jeder Spot entspricht einem anderen Peptid. Die einzelnen Membranen wurden je mit einem anderen Serum inkubiert; die oberen mit Seren jeweils einer weissen (BALB/c) und die unteren mit Seren jeweils einer schwarzen (C57BL/6) Maus. Die Abbildungen sind Negative der aus dem LumiImager<sup>TM</sup> stammenden Bilder; je dunkler der Spot desto mehr Signale der Enzymreaktion, desto mehr Antikörperbindungen.

(b) eines der schlechteren Beispiele (die noch auswertbar waren). Die mittleren Signal-Rausch-Verhältnisse (SNR, siehe Abschnitt 3.6.1) waren 1.71, bzw. 1.21 (entsprechen mittleren SNR innerhalb der Träger). Die Signale der Abbildung 4.7(a) nehmen von links nach rechts ab. Derartige Gradienten entstanden vermutlich durch die Abnahme der reaktiven Peptide in Aufdruck-Richtung auf die Glasträger oder resultierten von den inhomogenen Oberflächen der Glasträger.

Ein weiterer optischer Vergleich der Qualitätsverbesserung der Glasträger bietet Abbildung 4.8. Auf beiden Subarrays von unterschiedlicher Qualität (weil zu unterschiedlichen Zeiten hergestellt) wurde die gleiche Peptidbibliothek aufgetragen. Der Vollständigkeit wegen sollte erwähnt werden, dass es schon in frühen Stadien einzelne klare Bilder gab, die jedoch Einzelfälle blie-

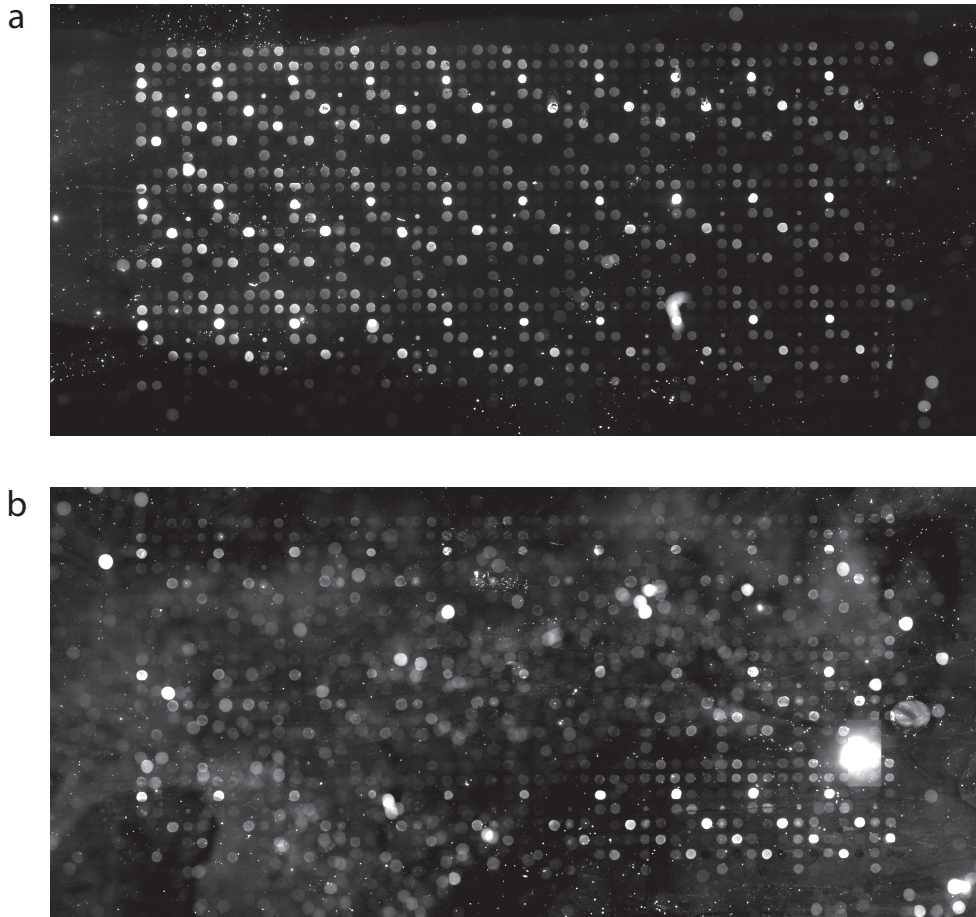


**Abbildung 4.6:** Die Verteilung der IgM-Signale auf einer Zellulosemembran. (a) Serumantikörperbindungen auf einer Zellulosemembran mit einer Bibliothek aus 508 Peptiden. Jeder Spot umfasst ein unterschiedliches Peptid. Die Bindungen wurden mittels Peroxidase-markierten sekundären Antikörpern detektiert. Je dunkler der Spot, desto mehr IgM ist gebunden. 440 Spots haben unterschiedliche Sequenzen und 34 Sequenzen sind dreimal wiederholt, (b) die Verteilung der 508 SI von IgM Peptidbindungsreaktivitäten in 90 Bins.

ben. Die variable, instabile Qualität der Inkubationsergebnisse rührte wohl von der Variabilität der Glasträgeroberflächen und deren Beschichtungen her.

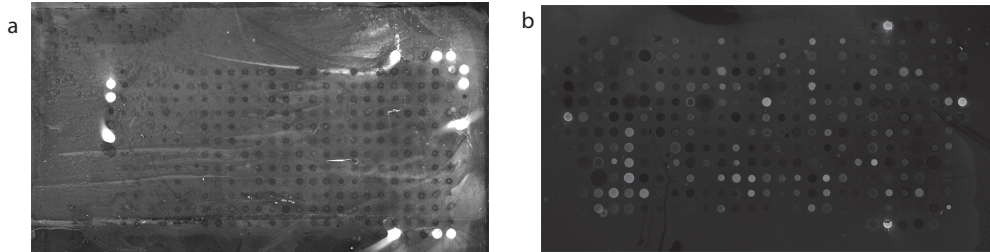
#### 4.4.2 Die Reproduzierbarkeit qualitativ hochwertiger Glasträger

Die Signale der Seruminkubationen auf den Chargen 879 und 901 zeigten eine sprunghafte Verbesserung der Signalqualität. Von gleichbleibend guten Trägeroberflächen resultierten erstmals auswertbare Signale, mit denen wir die Analyse der Serumantikörper-Reaktivitäts-Profile vornehmen konnten (siehe Kapitel 6). Die Abbildung 4.9 zeigt Mittelwerte  $\mu$  und Standardabweichung  $\sigma$  von 15 Repetitionen (von drei Trägern mit je fünf Subarrays) der Inkubationen des gleichen BALB/c Serum. Die 15 Repetitionen haben hohe mittlere Signal-Rausch-Verhältnisse:  $SNR(IgM) = 2.32$  und  $SNR(IgG) = 2.43$ .

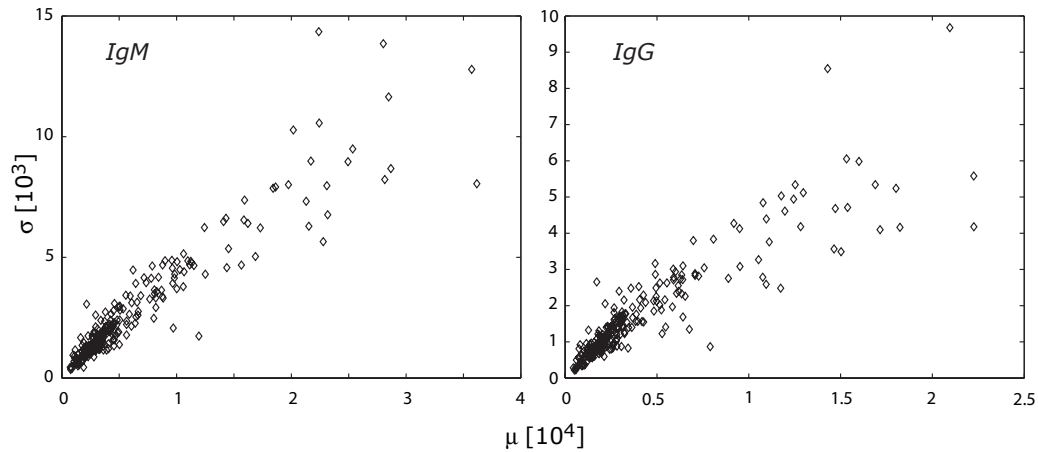


**Abbildung 4.7:** Peptidarrays auf Glasträgern mit vielen Repetitionen und von unterschiedlicher Qualität. (a) IgM-Ak-Signale eines BALB/c-Serums auf einem Glasträger der 380er Charge. (b) IgM-Ak-Signale eines anderen BALB/c-Serums auf einem Glasträger der 380er Charge. Die Glasträger haben die gleichen 40 Peptide in 30 Wiederholungen aufgetragen.





**Abbildung 4.8:** Die Qualitätszunahme der Inkubation von Seren auf Glasträgern anhand von zwei Beispiel-Subarrays. (a) Der Subarray stammt von einem Glasträger der 813er Charge und zeigt vor allem verwischte Kontrollen und kaum 'positive' Signale von bindenden Serumantikörpern. (b) Der Subarray stammt von einem Chip der 901er Charge und zeigt ein schönes Antikörper-Reaktivitäts-Profil anhand verschieden stark leuchtender Peptidspots.



**Abbildung 4.9:** Mittelwerte  $\mu$  und Standardabweichungen  $\sigma$  der SI von 15 Inkubations-Repetitionen auf Glasträgern. Die IgM-SI-Repetitionen links haben einen mittleren SNR=2.32, die IgG-SI-Repetitionen rechts einen mittleren SNR=2.43.

### 4.4.3 Signalintensitäten-Verteilungen der Glasträgerdaten

Repräsentative Beispiele der IgM- und IgG-Antikörpersignale aus jeder der vier Gruppen sind in Abbildung 4.10(a) auf acht Subarrays gezeigt. Die Verteilung der Daten lässt sich am schönsten mit vielen Repetitionen darstellen. Abbildung 4.10(b) zeigt 15 Repetitionen (von drei Trägern mit je fünf Subarrays) von Inkubationen des gleichen Serums ( $15 \times 252$  Datenpunkte) in einem Histogramm mit 50 Bins. Die Verteilung der Signalintensitäten entspricht einer Lognormal-Verteilung. Abbildung 4.11 zeigt eine logarithmierte Darstellung der 15 Subarrays, die mit demselben BALB/c Serum inkubiert worden sind, inklusive einer Anpassung (engl. *fit*) der Gaussverteilung mit  $\mu = 0.695$  und  $\sigma = 9.076$ .

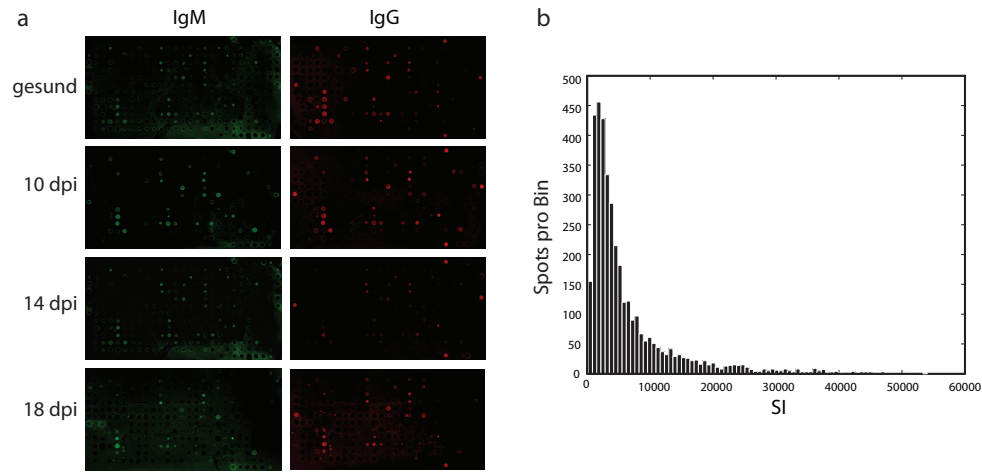
Eine Streuung der Signale zwischen einzelnen Subarrays existiert auch auf den 'guten' Glasträgern, wie die Boxplot-Abbildung 4.12 zeigt. Die Tatsache, dass derart grosse Varianzen eine signifikante und hohe diagnostische Zuverlässigkeit (DZ) (siehe Tabelle 6.1) *ohne* Normierung erlauben, spricht für eine grosse Robustheit der erzielten Ergebnisse (siehe Kapitel 6).

## 4.5 Auslesen der Signalintensitäten

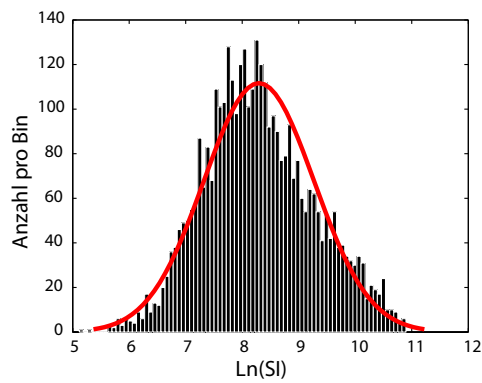
Das Auslesen der Signalintensitäten ist ein entscheidender Bestandteil der Datenanalyse. Werden die falschen Stellen auf den Trägern als Spot-Signale ausgelesen, sind die Daten unbrauchbar. Ebenso haben die Grössen der Radien der ausgelesenen Flächen wie auch die Verarbeitung der Hintergrund-Signale einen grossen Einfluss auf die Qualität der Resultate.

### 4.5.1 Software zum Auslesen der Signalintensitäten

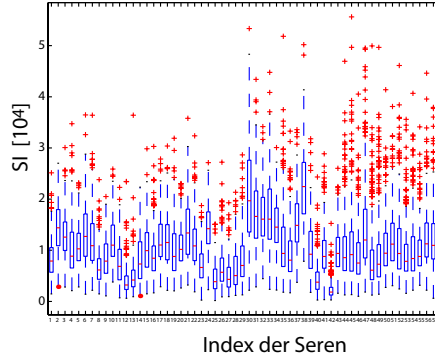
Die Ermittlung des idealen Programms zum Auslesen der Signalintensitäten ist kein abschliessend gelöstes Problem. Die Kosten der Programme sind hoch, weswegen nur diejenigen getestet wurden, die uns von Kooperationspartnern empfohlen oder zum Ausprobieren zur Verfügung gestellt wurden. Die Qualität der Software und ihre Brauchbarkeit für uns hängt einerseits vom Algorithmus zur Detektion der Spots (Auflegen des Gitters) und andererseits von den Möglichkeiten zur manuellen Gitterkorrektur ab. Diese Software-Funktionen hängen wiederum sehr vom Kontrast der Spotsignale gegenüber des Hintergrundes ab. Für Zellulosemembranen und Glasträger,



**Abbildung 4.10:** Signalintensitäten-Verteilung auf Glasträgern. (a) Signale von Serumantikörpern auf verschiedenen Subarrays verschiedener Peptidarray-Glasträger. Bindungen wurden mittels Fluoreszenz-markierter sekundärer Antikörper detektiert (grün: anti-IgM-Alexa Fluor 546, rot: anti-IgG-Alexa Fluor 647). Die Kontrollen in den Ecken waren TAMRA, eine Fluoreszenzkontrolle sowie murine IgM- und IgG-Antikörper (jeweils oben rechts und links und unten rechts). (b) Die Verteilung von  $15 \times 252$  IgM-Signalintensitäten von fünfzehn Inkubations-Repetitionen des gleichen BALB/c-Serums in 50 Bins.



**Abbildung 4.11:** Logarithmierte Signalintensitäten von 15 Inkubations-Repetitionen und deren Gaussverteilung: Lognormal der SI aus Abbildung 4.10 als Histogramm in 80 Bins verteilt. Eine Normalverteilung lässt sich gut an die empirische Verteilung anpassen (mit den Parametern  $\sigma = 9.076$ ,  $\mu = 0.695$ ).



**Abbildung 4.12:** Signalintensitäten-Verteilung aller BALB/c-Inkubationen auf Glasträgern mittels eines Boxplots der SI von jeder der 58 Inkubationen. Parallel nebeneinander aufgetragen sind die nicht normierten Mittelwerte, Standardabweichungen (blau) und Ausreisser (rot) von mit *H.-polygyrus*-infizierten (Index 1 – 28 früh nach Infektion, Index 29 – 43 spät nach Infektion) und gesunden BALB/c-Mäusen (Index 44 – 58).

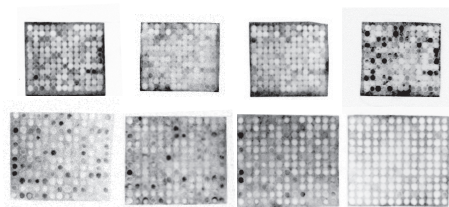
bei denen Herstellung oder Inkubation nicht von höchster Qualität waren, gibt es keinen Algorithmus der ein Gitter zuverlässig automatisiert findet. Für qualitativ hochwertige Bilder mit wenig Hintergrundrauschen hat sich *GenePix Pro 6 Microarray Image Analysis* von Molecular Device, CA, USA als bestes Programm erwiesen. *GeneSpotter<sup>TM</sup>* von MicroDiscovery GmbH Berlin hat sich als sehr geeignet zum Finden von Gittern auf Zellulosemembranen erwiesen. Es eignet sich auch hervorragend zum manuellen Auflegen und/oder Anpassen der Gitter. Eine detailliertere Beschreibung der verwendeten Software findet sich in Abschnitt 3.6.

#### 4.5.2 Signalintensitäten von Peptidspots und Hintergrund im Vergleich

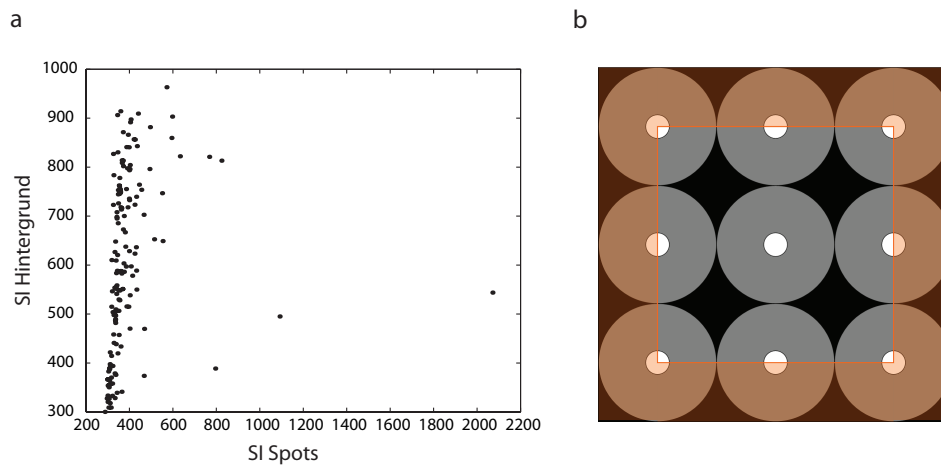
Zur Untersuchung des Einflusses der Hintergrundsignale wurden die Peptidspot-Signale und die Hintergrund-Signale einer repräsentativen Zellulosemembran miteinander verglichen. Entscheidend für die Bestimmung des Einflusses des Hintergrundes auf das eigentliche Signal ist nicht nur die effektive Technologie, sondern auch der Algorithmus, der zum Auslesen und Berechnen der Hintergrundsignale angewandt wurde. Das Auslesen der SI geschah mit zwei verschiedenen Softwarepaketen: *GeneSpotter* und *GenePix* (siehe auch

Abschnitt 3.6). *GeneSpotter* löst das Problem der Trennung der Bereiche des Peptidspots und des Hintergrundes elegant. Abbildung 4.14(b) zeigt, wie sich die Spot-SI (weiss) vom Hintergrund-SI (schwarz) trennt. Wir folgern, dass die Gefahr der Interferenz zwischen Peptidspot und Hintergrund abnimmt, je kleiner die Spot-Radien und je grösser die Abstände zwischen den Peptidspots werden. Ein repräsentatives Beispiel der Korrelationen zwischen derart ausgelesenen Signalen von Peptidspots und Hintergründen einer der vollautomatisch hergestellten Membranen aus Charge fünf zeigt Abbildung 4.14(a). Zwischen den SI der Peptidspots und den SI der Hintergrundsignale bestand eine schwache Korrelation ( $< 0.2$ ), die jedoch zu vernachlässigen war.

Darüber hinaus zeigen bereits die Bilder der inkubierten Membranen aus Charge fünf in Abbildung 4.13, dass die Signale des Hintergrundes von vollautomatisch hergestellten Membranen stärker schwanken als die Signale ihrer Peptidspots. Die Hintergrundsignale der Glasträger sind oft ebenfalls stärker verrauscht als die Peptidspots (Abbildungen 4.7(b) und 4.8(a)). Aus diesen Gründen verzichteten wir generell auf eine Korrektur der SI der Peptidspots mit den Hintergrund-Signalen.



**Abbildung 4.13:** Die Signale der Seruminkubationen gegenüber je 156 Peptiden auf acht Zellulosemembranen aus der manuell und vollautomatisch hergestellten Charge(n) fünf. Die oberen vier Membranen wurden manuell und die unteren vier vollautomatisch hergestellt. Die vier Zellulosemembranen auf der linken Seite wurden mit je einem verschiedenen C57BL/6-Serum und die vier auf der rechten Seite mit verschiedenen BALB/c-Seren inkubiert.



**Abbildung 4.14:** Die Abhängigkeit der Signalintensitäten der Peptidspots von den Hintergrundsignalen. (a) Die Signale der Peptidspots sind gegen die Signale des zugehörigen Hintergrundes von jedem der 156 Peptide einer voll-automatisch hergestellten, Serum-inkubierten Zellulosemembran aus Charge fünf aufgetragen. (b) Schematische Darstellung der Flächen zur Bestimmung der SI von Peptidspots und der SI des Hintergrundes mittels *GeneSpotter*. Bereiche, die für die Bestimmung der SI der Spots ausgelesen werden, sind Weiss, Bereiche für die Bestimmung des Hintergrundes sind Schwarz. Die graue Zone dazwischen ist ein Sicherheitsabstand.

### 4.5.3 Korrektur der sekundären Antikörper-Signale

Inkubationen auf Zellulosemembranen mit dem sekundären Antikörper alleine brachten keine bis vernachlässigbar schwache Signale. Deswegen wurde von einer Korrektur der Bindung der sekundären Antikörper an Peptidspots auf Zellulose abgesehen. Die Signale der sekundären Antikörper auf Glas-trägern waren vergleichsweise höher. Des Weiteren gab es einen unerklärten Additionseffekt: Peptide die von den sekundären Antikörpern alleine gebunden wurden, erhielten bei den Messungen der Bindungen von Seren, die mit dem jeweiligen Antikörper nachgewiesen wurden, um vielfach erhöhte SI. Aus diesem Grund wurden die Peptide mit höchsten SI der sekundären Antikörper alleine, vor der Analyse aus allen Daten entfernt. Typischerweise waren das aber nur zwei bis drei (der 255) Peptide, die durch sekundäre Antikörper-Bindungs-Korrekturen elimiert wurden.

## 4.6 Peptid-Peptid Bindungsversuch auf Glasträgern

Um die Frage zu klären, ob es möglich ist, Bindungen von Peptiden (statt Proteinen, wie z. B. Antikörpern) auf systematischen Peptidbibliotheken zu messen, haben wir folgenden Versuch durchgeführt: Der Versuch, ein einzelnes Peptid mit systematischen Peptidbibliotheken zu inkubieren und eine Bindung nachzuweisen, wurde mittels eines von JPT zur Verfügung gestellten Peptids mit Molekülmasse 3296.86 g/Mol durchgeführt. Das Samplepeptid hatte ein N-terminales Fluoreszenz-Label (5(6)-Carboxyfluorescein). Der Versuch, 1.6 mg des Samplepeptids in 500  $\mu$ l Inkubationsflüssigkeit ( $9.7 \times 10^{-4}$  mol Samplepeptid ) auf zwei verschiedenen Glasträgern mit Bibliotheken für Kinasen (Charge 285) und Rezeptor-Scans (Charge 309) im Sandwichversuch (d.h. ein Mikroarray als Träger, ein Mikroarray als Deckel) zu inkubieren, erbrachte keine nachweisbaren Signale, bzw. keine nachweisebare Bindung des Samplepeptids an eines der Probepeptide der Träger.

Das experimentelle Ergebnis, dass derartige Bindungen, wohl aufgrund der Flexibilität aller Bindungspartner, nicht messbar sind, ist nicht unerwartet.





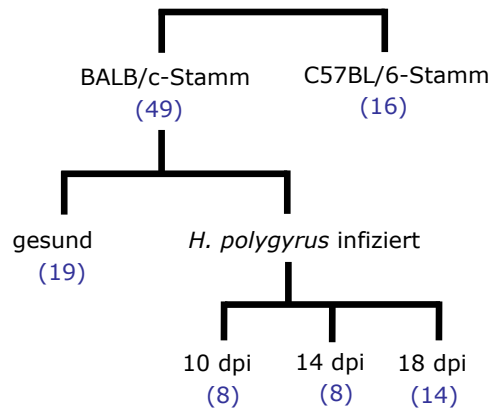
# Kapitel 5

## Resultate Zellulosemembran

In diesem Kapitel werden die IgM-Serumantikörper-Reaktivitäts-Profile der untersuchten Samples besprochen. Wie in Kapitel 4 beschrieben, ist die Reproduzierbarkeit der Signale auf Zellulosemembranen eher gering. Wir zeigen, dass wir dennoch Gruppen von Seren verschiedener Mäuse klassifizieren können. Die analysierten Serensamples und ihre Gruppenzugehörigkeit sind in Abbildung 5.1 dargestellt. Jedes der 65 Serensamples wurde auf einer separaten Zellulosemembran inkubiert. Die Bindungen der IgM-Serumantikörper an die einzelnen Peptid-Spots wurden entsprechend der Beschreibung in Abschnitt 3.4 quantifiziert. Die quantifizierten Bindungen wurden in Signalintensitäten (SI) ausgedrückt. Die 156 relevanten SI jeder Inkubation wurden zur Datenanalyse als einzelne Vektoren in die Datenmatrix übertragen.

### 5.1 Hauptkomponentenanalyse und Normierung der Daten

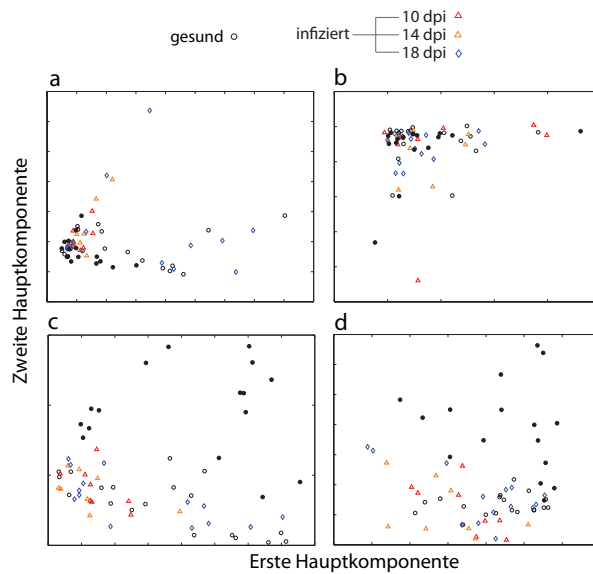
Die unterschiedlichen Möglichkeiten der Normierung des Membrandatensatzes wurden mittels Hauptkomponentenanalyse (siehe Abschnitt 3.9.1) analysiert. Mit der Hauptkomponentenanalyse (PCA) werden die Daten einer Matrix nach Grösse ihrer Varianzen (Hauptkomponenten) sortiert (siehe Kapitel 1.7). Die grafische Darstellung der ersten Hauptkomponenten visualisiert die Verteilung der Vektoren (Samples) einer Matrix in übersichtlicher Weise. Die PCA-Abbildungen 5.2(a)-(d) zeigen, weshalb die SI der Inkubationen der Serensamples auf Zellulosemembranen mit der Formel  $\|nv_i\| = 1$  (Formel 3.5) normiert wurden. Die ersten beiden Hauptkomponenten der PCA des Datensatzes der Zellulosemembran, der nach den in Abschnitt 3.7 beschriebenen Formeln normiert wurde, zeigen:



**Abbildung 5.1:** Die Mäusegruppen, deren Seren auf Zellulosemembranen inkubiert und analysiert wurden. 65 Serensamples von BALB/c- und C57BL/6-Mäusen wurden untersucht. Von den 65 Samples stammen 35 von gesunden Mäusen und 30 von *H.-polygyrus*-infizierten. Die *H.-polygyrus*-infizierten trennen sich in 8 Serensamples von 10 Tagen nach Infektion (10 dpi), 8 von 14 Tagen nach Infektion (14 dpi) und 14 von 18 Tagen nach Infektion (18 dpi) auf.

- Ohne Normierung streuen die Daten unregelmässig (Abbildung 5.2(a)).
- Die Subtraktion des Medians (Formel 3.3) bringt die blauen Rauten (Gruppe »spät nach Infektion«) näher zusammen. Deren Streuverhalten wird in Abschnitt 6.5 beschrieben. Abbildung 5.2(b) zeigt aber auch, dass sich dadurch an der Verteilung der Daten sonst wenig ändert.
- Die Median-Standardabweichung normierter Daten nach Formel 3.4 in Abbildung 5.2(c) separiert vor allem die C57BL/6-Samples untereinander. Diese werden so aber auch verstärkt nach ihrer Herkunft und ihren Inkubationszeitpunkten aufgetrennt.
- Unter Verwendung der Formel 3.5 werden die verschiedenen Stämme aufgetrennt, und die Daten gleichmässig verteilt.

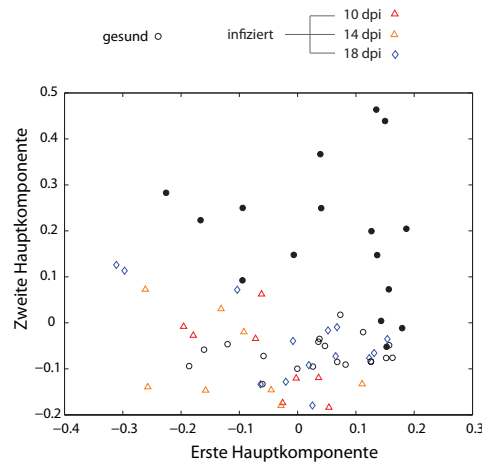
Die Abbildungen 5.2(a)-(d) zeigen, dass die Normierung nach Formel 3.5 die Daten in idealer Weise verteilt und nach Stämmen aufteilt.



**Abbildung 5.2:** Hauptkomponentenanalysen und Normierung der Inkubationen von Seren auf Zellulosemembranen. Die ersten zwei Hauptkomponenten von unterschiedlich normierten Mäuseseren-SI auf Zelluloseträgern. (a) zeigt die Hauptkomponentenanalyse der nicht normierten Daten, (b) jene der nach Formel 3.3 normierten, (c) der nach Formel 3.4 normierten und (d) der nach Formel 3.5 normierten Daten. Die Beschriftung der Größenordnung der Achsen ist weggelassen; das Ziel dieser Abbildungen ist eine Veranschaulichung der Verteilungen der Serensamples in der Ebene aufgrund der Normierung. Die vier inkubierten Sample-Gruppen sind: gesunde (o) BALB/c-Mäuse und infizierte BALB/c-Mäuse 10 ( $\Delta$ ), 14 ( $\triangle$ ) and 18 ( $\diamond$ ) Tage nach Infektion. Des Weiteren ist PCA ein unvoreingenommener Algorithmus und die Labels der Datenpunkte wurden nur zur Visualisierung nachträglich angebracht.

## 5.2 Grenzen der Hauptkomponentenanalyse

Die ersten beiden Hauptkomponenten der Hauptkomponentenanalyse (PCA) der normierten Signalintensitäten des Zellulosemembran-Datensatzes trennen die Seren-Datenpunkte der unterschiedlichen Stämme bereits auf, siehe Abbildung 5.3. Mäuse aus verschiedenen Stämmen lassen sich also aufgrund der grössten Varianzen in den Daten separieren. Eine Separation der gesunden und kranken Mäuse innerhalb eines Stammes ist jedoch mit PCA nicht möglich. Dafür bietet sich die Lineare Diskriminanzanalyse (LDA) auf einer dimensionsreduzierten Matrix an.



**Abbildung 5.3:** Die ersten zwei Hauptkomponenten von Mäuseseren-Ak-Reaktivitäten auf Zellulosemembranen. Jeder Datenpunkt entspricht den normierten SI der Antikörperbindungen eines inkubierten Serums. Die Serumdatenpunkte wiederum entsprechen Vektoren mit 156 verschiedenen SI, die von den 156 Peptidspots stammen, die auf allen analysierten Zellulosemembranen vorhanden sind. Die erste Komponente enthält 41.1% der Varianz, die zweite 13.1%. Die Trennung der beiden Stämme ist bereits mit der erwartungstreuen Hauptkomponentenanalyse sichtbar.

## 5.3 LDA, Dimensionsreduktion und Merkmalsselektion

Die Berechnung der Linearen Diskriminanzanalyse (LDA) erfordert eine Inversion der Datenmatrix. Die normierten SI der Bindungsmessungen des Zellulosemembrandatensatzes korrelieren hoch. Hoch-korrelierende Matrizen können nicht invertiert werden. Deshalb ist die Berechnung der LDA nur auf einem dimensionsreduzierten Datensatz möglich. Dimensionsreduktion kann durch willkürliche Datenelimination oder durch Merkmalsselektion erfolgen, wie in Kapitel 1.7 beschrieben. Reduzierte Datensätze mit minimalem Varianzverlust erhält man mittels PCA. Durch die Wahl der Zahl der Hauptkomponenten kann sowohl die Zahl der Dimensionen als auch der Varianz des reduzierten Datensatzes beliebig variiert werden (siehe Abschnitt 1.7.1). Um beispielsweise die 65 Seren vollständig in BALB/c- und C57BL/6-Klassen zu klassifizieren, benötigt man 14 Hauptkomponenten (Dimensionen), wie in Abbildung 5.4(a) gezeigt wird. Eine Klassifizierung der gesunden und infizierten Seren innerhalb des Stammes ist jedoch unabhängig von der gewähl-

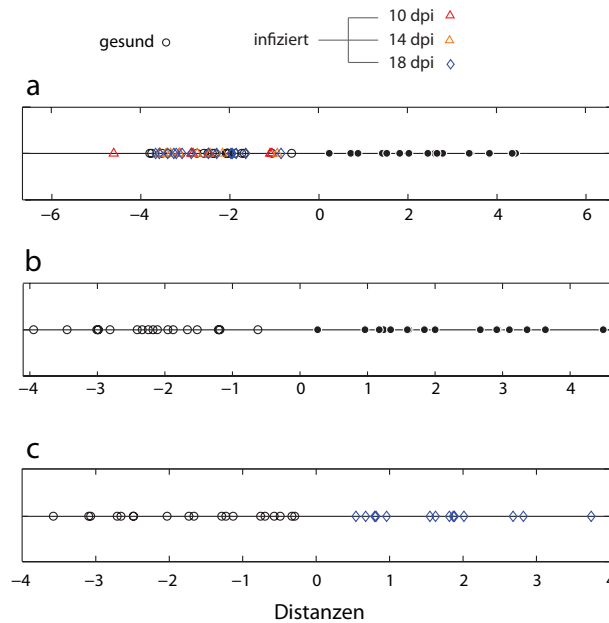
ten Zahl von Hauptkomponenten mit LDA machbar. Um diese vollständig zu klassifizieren wird eine Klassifizierungsergebnis-optimierte Dimensionsreduktion benötigt. Die selektierten Dimensionen (= Merkmale = Peptide) werden dazu mittels Merkmalsselektion (siehe Abschnitt 1.7.2) 'herausgefischt'.

Die Merkmale werden mittels des Algorithmus der Potential Support Vector Machine (P-SVM) selektiert. P-SVM ist in Abschnitt 3.9.9 beschrieben und hat sich als für unsere Daten sehr geeigneter Algorithmus erwiesen. Optimalerweise konnten wir darüber hinaus die Einstellungen und Eigenschaften von P-SVM unseren Bedürfnissen und derer unserer Daten anpassen lassen.

Mit nur acht P-SVM-selektierten Peptiden lassen sich beispielsweise die 19 gesunden BALB/c-Datenpunkte und die Datenpunkte der 14 *H.-polygyrus*-infizierten BALB/c-Seren 18 Tage nach Infektion (dpi, *days post infection*) vollständig klassifizieren (33 Datenpunkte in acht Dimensionen), wie Abbildung 5.4(c) zeigt. Durch Merkmalsselektion lassen sich nicht nur schwer unterscheidbare Daten klassifizieren. Auch die Zahl der benötigten Dimensionen (Merkmale) wird klein gehalten. Die 14 Dimensionen zur vollständigen Klassifizierung aller 65 Seren in BALB/c- und C57BL/6-Klassen (Abbildung 5.4(a)) lassen sich mittels P-SVM-Merkmalsselektion auf fünf »beste« Dimensionen reduzieren. Die gesunden BALB/c-Seren lassen sich mit fünf ausgewählten Peptiden gegenüber den gesunden C57BL/6-Seren vollständig klassifizieren (Abbildung 5.4(b)).

LDA ist ein voreingenommener Algorithmus, der sich gut zur Visualisierung von klassifizierten Daten eignet. Entscheidend für die Qualität der Resultate sind die Anzahl und Auswahl benötigter Dimensionen, die an LDA übergeben werden.

Die P-SVM erlaubt nebst der Merkmalsselektion auch die Berechnung der Präzision der Klassifizierung und der Vorhersagesicherheit (mittels 'leave one out'-Kreuzvalidierung) aufgrund der ausgewählten Merkmale. Selektion und Präzisionsberechnungen sind jedoch nur auf zweigeteilten Datensätzen möglich. Tabelle 5.1 zeigt die Klassifizierungs- und Vorhersagepräzisionen zwischen der Gruppe der gesunden BALB/c-Seren und allen anderen untersuchten Gruppen. Die Anzahl der benötigten Merkmale und mit Labelaustausch berechneten Signifikanzen (Abschnitt 3.9.10) sind dort ebenfalls angegeben. Klassifizierungspräzisionen sind bei allen Gruppen über 96%. Die Kreuzvalidierungs-Vorhersagepräzisionen sind über 75%. Die Signifikanzen der Resultate sind allerdings nur zwischen den zwei Mäusestämmen und zwischen den gesunden und *H.-polygyrus*-infizierten 18 dpi BALB/c-Seren



**Abbildung 5.4:** LDA Klassifizierung des Zellulosemembrandatensatzes – mit und ohne Merkmalsselektion. (a) LDA der ersten 14 Hauptkomponenten (91.17% der Varianz) der normierten SI der gesunden und infizierten BALB/c-Seren gegenüber der gesunden C57BL/6-Seren (65 Datenpunkte in 14 Dimensionen). (b) LDA der normierten SI von fünf ausgewählten Peptiden trennt die gesunden 19 BALB/c- und 16 C57BL/6-Seren (35 Datenpunkte in fünf Dimensionen). (c) LDA der normierten SI von acht selektierten Peptiden trennt die 19 gesunden BALB/c-Seren und die 14 *H. polygyrus*-infizierten BALB/c-Seren 18 dpi (33 Datenpunkte in acht Dimensionen).

befriedigend. Die schlechten Signifikanzen zwischen den gesunden und infizierten BALB/c-Seren von früheren Zeitpunkten lassen sich aufgrund der Daten schlecht erklären, doch könnten die kleinen Gruppengrößen der infizierten von früheren Zeitpunkten zumindest ihren Teil beitragen.

Die Resultate sind aber vor allem angesichts der Reproduzierbarkeit der Technologie (Abschnitt 4.3) von erstaunlich guter Qualität.

## 5.4 LDA-Multiklassen-Klassifizierung

P-SVM erlaubt nur Klassifikationen und Vorhersagen von zwei Gruppen untereinander. Die gewählten Merkmale zwischen den Gruppenpaaren können aber als größeren Datensatz an LDA übergeben werden, um eine Multiklassen-

Klassifizierung mehrerer Gruppen zu erhalten. Abbildung 5.5 zeigt eine zwei-dimensionale LDA-Klassifizierung von vier Gruppen der gesunden und infizierten BALB/c-Mäuseseren der verschiedenen Zeitpunkte. Der zeitliche Verlauf der Infektion zeigt sich durch die Auftrennung der vier Gruppen im Gegenurzeigersinn. Dazu wurde LDA auf einem reduzierten 22-dimensionalen Datensatz berechnet, der aus ausgewählten Peptiden bestand. Die 22 Peptide (Dimensionen) des reduzierten Datensatzes wurden mittels P-SVM-Klassifizierung zwischen allen BALB/c-Gruppen selektiert. Die C57BL/6-Gruppe wurde weggelassen, weil sie sich von allen andern sehr einfach trennt und das Bild nur unübersichtlicher macht.

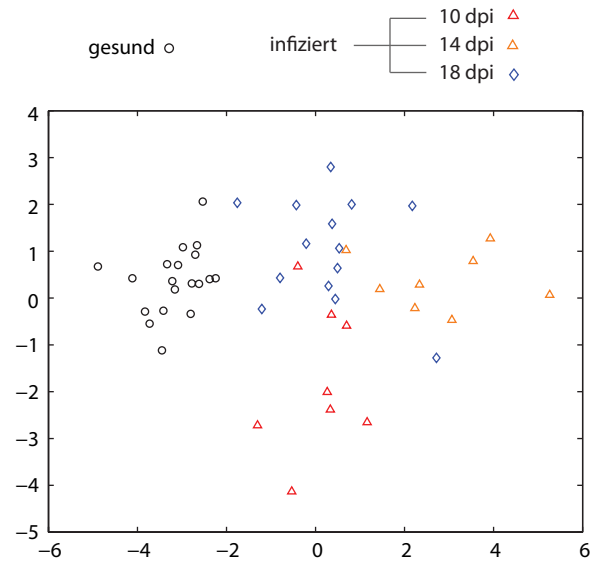
LDA-Multiklassen-Klassifizierung in vier Gruppen aufgrund von 22 Merkmalen klassifiziert nicht fehlerfrei aber anschaulich.

	gesunde BALB/c versus gesunde C57BL/6	gesunde BALB/c versus infizierte 10 dpi	gesunde BALB/c versus infizierte 14 dpi	gesunde BALB/c versus infizierte 18 dpi
Klassifizierungspräzision (%)	100	96.3	100	100
Anzahl Merkmale	4	5	5	7
Signifikanz $p$	0.001	0.25	0.16	0.06
Vorhersagepräzision (%)	97.14	77.78	81.42	75.76
Anzahl Merkmale	4	6	9	3
Signifikanz $p$	0.001	0.132	0.246	0.028

**Tabelle 5.1:** Klassifizierung und Kreuzvalidierungs-Vorhersage der gesunden BALB/c- gegenüber der C57BL/6-Mäusen und der *H.-polygyrus*-infizierten-BALB/c-Mäusen. P-SVM Merkmalsselektion und ihre Zuverlässigkeitsparameter: Die Klassifizierungen und Kreuzvalidierungs-Vorhersagen der IgM-Bindungsreaktivitäten für alle Gruppenpaare, und die mittels Labelaustausch berechneten Signifikanzen  $p$ .

## 5.5 Merkmalsselektion versus Normierung

Eine eingehendere Untersuchung der Peptide, die sich für eine gute Klassifizierung oder Vorhersage von zwei Gruppen eignen, hat für die selektierten Peptide der Zellulosemembranendaten wenig Sinn gemacht. Alle Charakterisierungen von Peptiden, die sich gut zur Klassifizierung eignen, sind auf eine Normierung angewiesen. Normierung erfolgte aber über alle Peptide einer



**Abbildung 5.5:** Zweidimensionale LDA-Klassifizierung aller vier BALB/c-Gruppen der Zellulosemembrandaten. Der zeitliche Verlauf der Infektion zeigt sich durch die Auftrennung der vier Gruppen im Gegenuhrzeigersinn. LDA wurde auf einem reduzierten 22-dimensionalen Datensatz, der aus ausgewählten Peptiden bestand, berechnet. Die 22 Peptide wurden mittels LDA-Klassifizierung zwischen allen Gruppen selektiert.

Membran. Unabhängige Betrachtungen der einzelnen Peptide und von kleineren Peptidgruppen sind deshalb gar nicht möglich. Weitere Überlegungen zu den Peptid- oder Merkmalsgruppen folgen in Abschnitt 6.5.

Zusammenfassend können wir festhalten, dass ein Vergleich der Bindungsmessungen von Serumantikörper-Reaktivitäts-Profilen gegenüber Zufallsbibliotheken von Peptiden auf Zellulosemembranen, verschiedene Immunzustände voneinander unterscheiden kann. Allerdings sind Unterscheidungen nur mit grossem methodischen Aufwand auffindbar und lassen ausserdem in ihrer Signifikanz oft zu wünschen übrig.



# Kapitel 6

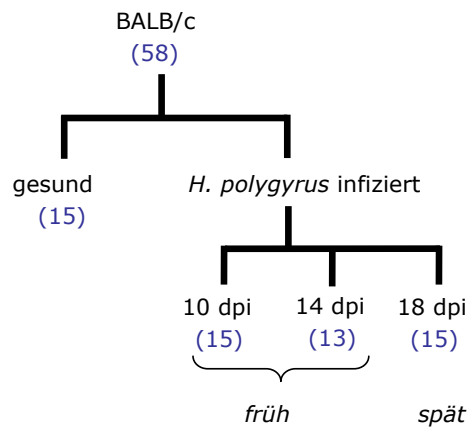
## Resultate Glasträger

In Kapitel 4.2 hatten wir beschrieben, dass die Bindungen der Serumantikörper auf systematischen Peptidbibliotheken auf Zellulosemembranen mehr als 10% nachweisbare Signale ergeben. Anzahl und Qualität der Signale von Serumantikörperbindungen auf systematischen Bibliotheken auf Glasträgern stellten uns vor neue und unbekannte Fragestellungen. Die ersten Experimente ergaben noch keine vielversprechend Resultate (Abschnitt 4.4.1). Zahlreiche Versuchsoptimierungen und wegweisende Fortschritte in der Herstellung der Glasträger führten schliesslich zu erstaunlichen Ergebnissen (4.4.2).

Die auf Glasträger untersuchten Seren stammten von den gleichen BALB/c-Mäusen, mit und ohne *H.-polygyrus*-Infektion, wie diejenigen, die, wie in Kapitel 5 beschreiben, auf Zellulose inkubiert und ausgewertet worden waren. Auf die Analyse der Inkubationen der C57BL/6-Seren wurde verzichtet, weil sie, wie bereits im vorhergehenden Kapitel gezeigt, sehr gut von den gesunden und infizierten BALB/c klassifizierbar sind. Resultate der Klassifizierung von BALB/c- vs. C57BL/6-Seren-Inkubationen auf Glasträgern werden in Bongartz u. a. [10] besprochen. Dieses Kapitel handelt ausschliesslich von der Analyse der diagnostischen Zuverlässigkeit (DZ) der *H.-polygyrus*-Infektionen und des Krankheitsverlaufes anhand von Serumantikörper-Reaktivitäts-Profilen. Untersucht wurden 58 Serensamples von BALB/c-Mäusen. Davon stammten 15 von gesunden und 43 von *H.-polygyrus*-infizierten zu unterschiedlichen Zeitpunkten nach Infektion, Abbildung 6.1.

### 6.1 Verzicht auf Normierung der Daten

Die Analyse der SI der Mäuseserum-Inkubationen auf Glasträgern konnte ohne Normierung erfolgen.



**Abbildung 6.1:** Mäusegruppen, deren Seren auf Glasträger inkubiert und analysiert wurden. 58 Serensamples von BALB/c-Mäusen wurden untersucht. Von den 58 Samples stammten 15 von gesunden Mäusen und 43 von *H.-polygyrus*-infizierten. Die *H.-polygyrus*-infizierten trennen sich in 15 Serensamples von 10 Tagen nach Infektion (10 dpi), 13 von 14 Tagen nach Infektion (14 dpi) und 15 von 18 Tagen nach Infektion (18 dpi) auf. Aufgrund der möglichen und unmöglichen Klassifizierungen lassen sich die *H.-polygyrus*-infizierten auch in »früh« (28 Samples) und »spät« (15 Samples) unterteilen.

Auf eine Normierung der Datensätze der Glasträgerinkubationen konnte aus folgenden Gründen verzichtet werden:

- PCA (Abbildung 6.3) zeigt, dass schon die unnormierten Daten gut verteilt sind.
- P-SVM selektierte Merkmale normieren sich selbst (Abbildung 6.2), was Hochreiter u. Obermayer [25] in ihren Korrelationsüberlegungen vorhergesagt haben.
- keine Formel aus Abschnitt 3.7 brachte eine Verbesserung der Klassifizierbarkeit der Datensätze untereinander (siehe dazu auch Abschnitt 8.3).

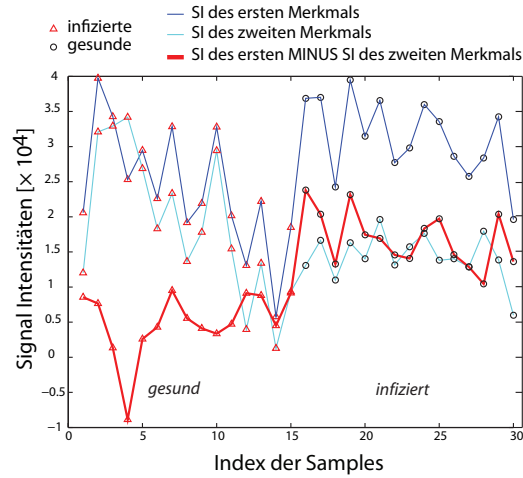
Die Potential Support Vector Machine (P-SVM), die in Abschnitt 3.9.9 beschrieben ist, eignet sich zum Auffinden selbstnormierender Merkmale. Abbildung 6.2 zeigt einerseits ein P-SVM-selektiertes Normierungsbeispiel und beweist des Weiteren die von Hochreiter u. Obermayer [25] vorhergesagten Korrelationsüberlegungen.

**Normierung** Die SI von zwei mittels P-SVM-selektierten Merkmalen (Beispiel Abbildung 6.2) normieren sich mittels Subtraktion selbst; die resultierenden Werte (rote Kurve) der linken 15 Samples (gesund) liegen deutlich tiefer ( $< 10^4$ ) als die der 15 infizierten Samples rechts ( $> 10^4$ ).

**Korrelationsüberlegung** Abbildung 6.2 beweist die Vorhersage in Hochreiter u. Obermayer [25] ( $t = f_1 + f_2$ , ohne dass  $t$  mit den Merkmalen  $f_1$  und  $f_2$  korrelieren muss). Der Vorteil von P-SVM gegenüber den meisten anderen Merkmalsselektions-Methoden ist, dass sie derartige Merkmalsgruppen auffinden kann.

## 6.2 Hauptkomponentenanalyse

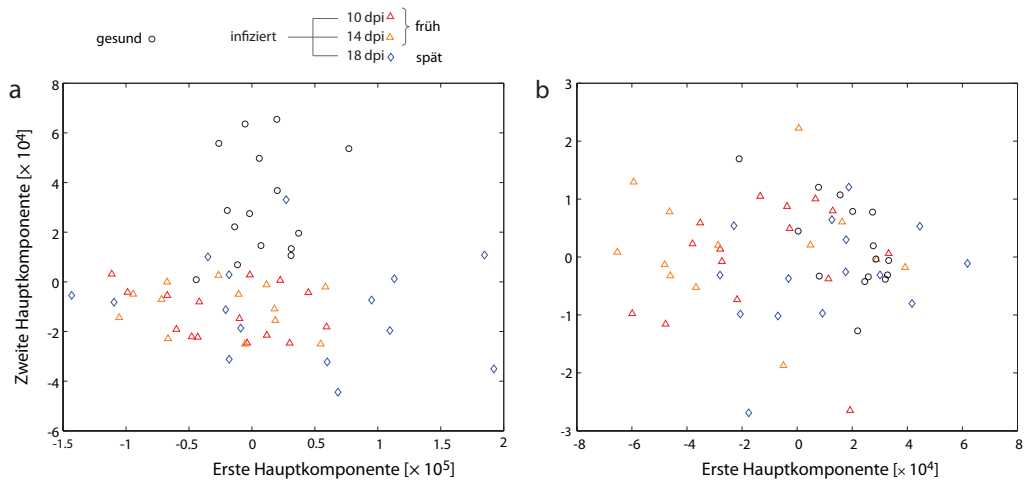
Die Hauptkomponentenanalyse (PCA) in Abbildung 6.3(a) zeigt, dass die IgM-Antikörpersignale der gesunden Seren von den *H.-polygyrus*-infizierten entlang der zweiten Hauptkomponente aufgetrennt werden. Die einzelnen Untergruppen der Seren von *H.-polygyrus*-infizierten BALB/c-Mäusen lassen sich mittels PCA nicht anhand der grössten Varianzen in den Daten auftrennen.



**Abbildung 6.2:** Normierung und Additionseffekte der P-SVM-selektierten Peptide. Die nicht normierten SI der zwei besten Merkmale für die Klassifizierung der Gruppe der gesunden (links) und der früh infizierten (rechts) Seren. Rot ist das Resultat der Subtraktion der beiden SI.

## 6.3 Reduktion der Zahl analysierter Gruppen

Mit P-SVM finden wir neun Merkmale, welche die infizierten Mäuse zwischen Tag 10 und Tag 14 zu 100%ige klassifizieren. Die Signifikanzberechnung dieser Klassifizierung mittels 500maligem Labeltausch (engl. *label shuffling*, siehe Abschnitt 3.9.10) ergibt, dass P-SVM auch 443 von 500 Mal eine 100%ige Klassifizierung der zufällig durchmischten Gruppen findet, was einer Signifikanz von  $p = 0.886$  entspricht (d.h. hochgradig NICHT signifikant). Die Wahl einer geringeren Klassifizierungsrate mit einer kleineren Anzahl Merkmale verhilft nicht zu einem sinnvolleren Ergebnis. Sieben Merkmale klassifizieren zu 92.8% mit einer noch schlechteren Signifikanz ( $p = 0.96$ ), und drei Merkmale klassifizieren nur noch zu 78.97% mit einer Signifikanz von lediglich  $p = 0.99$ . Durch Reduktion der Anzahl der Merkmale erzielt man also keine bessere Signifikanz. Die Bindungseigenschaften der Seren von 10 dpi und 14 dpi sind sich anscheinend zu ähnlich, um mit den Methoden, die in dieser Arbeit getestet wurden, eine statistisch relevante Klassifizierung zu erhalten. Andere getestete Klassifizierungsmethoden wie LDA finden auch keine Separation dieser beiden Gruppen (Daten nicht gezeigt). Die Seren der 10 dpi und 14 dpi *H.-polygyrus*-infizierten Mäuse werden daher in eine Gruppe von mit *H.-polygyrus*-infizierten Mäusen zu einem **frühen**



**Abbildung 6.3:** Hauptkomponentenanalyse (PCA) der IgM- und IgG-Antikörpersignale von auf Glasträgern inkubierten Seren. (a) Die ersten zwei Hauptkomponenten der IgM-Signale der Antikörper-Bindungen der Mäuseseren. Jeder Datenpunkt entspricht einem inkubierten Serumsample. Jedes Sample entspricht einem Vektor mit den SI von jedem der 252 Peptide. Die 252-dimensionalen Vektoren stammen von Inkubationen auf Glasträger-Subarrays, welche in 252 SI von ebenso vielen verschiedenen Peptidspots resultieren. (b) Die ersten zwei Hauptkomponenten der IgG-Signale der Antikörper-Bindungen der Mäuseseren. Die Daten entstanden wie bei (a), ausser dass bei den IgG-Antikörper-Analysen, aufgrund der Korrekturen der sekundären Antikörper-Inkubationen 253 Peptide (bzw. SI) pro Sample einbezogen wurden. Die vier inkubierten Sample-Gruppen sind: gesunde (o) BALB/c-Mäuse und infizierte BALB/c-Mäuse 10 ( $\Delta$ ), 14 ( $\Delta$ ) and 18 ( $\Delta$ ) Tage nach Infektion.

Zeitpunkt nach Infektion (Label »früh«) zusammengefasst. Den Seren der 18 dpi *H.-polygyrus*-infizierten Mäusen geben wir das Label »spät«). Durch das Zusammenlegen der frühen Zeitpunkte bleiben mit den Seren der gesunden BALB/c-Mäuse drei Gruppen, die sich zu vier Vergleichs-Paaren gruppieren lassen.

	gesund im Vergleich zu früh nach Infektion		gesund im Vergleich zu spät nach Infektion		früh im Vergleich zu spät nach Infektion	
	IgM	IgG	IgM	IgG	IgM	IgG
Sensitivität	1	1	0.93	0.80	0.93	0.86
Spezifizität	0.96	0.93	1	0.87	1	1
DZ (%)	97.7	90.7	96.7	83.4	95.3	90.7
# Merkmale	2	2	6	4	5	4
Signifikanz $p$	<0.002	<0.002	0.004	0.014	<0.002	<0.002

**Tabelle 6.1:** Sensitivität und Spezifizität der Klassifizierung der gesunden und *H.-polygyrus*-infizierten BALB/c-Mäusen der Glasträgerdatensätze. P-SVM-Merkmalss Selektion und ihre Zuverlässigkeitsparameter: Kreuzvalidierung-Vorhersage der IgM- und IgG-Bindungsreaktivitäten für alle klassifizierten Mäusegruppen-Paare und mittels Label-Austausch berechnete Signifikanzen ( $p$ ). DZ steht für Diagnostische Zuverlässigkeit.

## 6.4 Paarweise Klassifizierung von drei Gruppen mittels Merkmalsselektion

Die verbleibenden drei Gruppen der gesunden und »früh«- und »spät«-nach-Infektion BALB/c-Seren wurden alle untereinander verglichen. Tabelle 6.1 zeigt, dass sich die vier verbleibenden Gruppen aufgrund der P-SVM-selektierten Merkmale mit sehr hohen diagnostischen Zuverlässigkeiten (DZ) und Signifikanzen paarweise auftrennen liessen. Sensitivität und Spezifizität wurden mit den Formeln 3.6 und 3.7 in Abschnitt 3.9.5 berechnet. Die diagnostische Zuverlässigkeit (DZ) wurde mittels leave-one-out-Kreuzvalidierung-Vorhersage bestimmt. Die DZ der IgM-Ak-SI lag zwischen 95.3% und 97.7% und die der IgG-SI zwischen 83.4% und 90.7%; die Zuverlässigkeiten der Vorhersagen der IgM-Ak-SI übertrafen die der IgG-Ak-SI (siehe auch Abschnitt 8.7.3). Die hohen DZ stammten darüber hinaus von sehr kleinen Zahlen von Peptiden (d.h. Merkmalen); nur zwei bis sechs Merkmale waren zur Unterscheidung ausreichend. Tabelle 6.2 zeigt die Sequenzen der selektier-

ten Peptide. Die mittels Labelaustausch berechneten Signifikanzen sind alle  $p \ll 0.05$  und damit hochgradig signifikant. Diese Resultate übertrafen unsere Erwartung, da sie höhere diagnostische Zuverlässigkeiten (DZ) aufweisen als herkömmliche kommerzielle Diagnosetests.

	gesund im Vergleich zu früh nach Infektion	gesund im Vergleich zu spät nach Infektion	früh im Vergleich zu spät nach Infektion
	IgM	IgM	IgM
Merkmale	THEDFRYDDVFEGN FFDEIIHSCRSQNG	LLMKYAFDDRQAEW PDWLTILYSFWFES THEDFRYDDVFEGN QQAIGKETTIELD SVQFWFRQGREQTG NTMNWHLGCDPFL	VRQVQRSKMHKKG KELDLEHKPRLHIT FFDEIIHSCRSQNG HSDKTIKTYTKMLR QSFTQKNNAIQFNK
	IgG	IgG	IgG
Merkmale	QQAIGKETTIELD DVCSWTSSDMTNMG	LLMKYAFDDRQAEW PPNEWMDDWQPGEN LDEEVFTFENLRIY RFRGTCPMWKCHFL	TERIKSELLGQNTK VIYDMEYSECYSVG KELDLEHKPRLHIT VRQVQRSKMHKKG

**Tabelle 6.2:** Die Sequenzen der Detekton-Merkmale zwischen allen klassifizierten Mäusegruppen-Paaren auf Glasträgern. Nur wenige der Peptide kommen in mehr als einer Gruppe vor. Peptid THEDFRYDDVFEGN taucht in der Gruppe »gesund vs. früh nach Infektion« und in der Gruppe »gesund vs. spät nach Infektion« auf, und Peptid FFDEIIHSCRSQNG kommt in der Gruppe »gesund vs. früh nach Infektion« und in der Gruppe »früh vs. spät nach Infektion« vor. Auch zwischen IgM- und IgG-Ak-Gruppen gab es Gemeinsamkeiten: Peptid QQAIGKETTIELD kommt in unterschiedlichen Gruppen der beiden Ig-Ak-Klassen vor, und die Gruppen »gesund vs. spät nach Infektion« enthalten sowohl in der IgM- als auch in der IgG-Ak-Vorhersage das Peptid LLMKYAFDDRQAEW. Die Gruppe »früh vs. spät nach Infektion« weist in den beiden Ig-Ak-Klassen sogar zwei gemeinsame Peptide auf.

## 6.5 Vom Protekton zum Detekton

### 6.5.1 Definition Detekton

In Anlehnung an die Definition des Protektions (Abschnitt 1.2.7) haben wir den Begriff des »Detektions« eingeführt. Detekton stützt sich dabei auf die

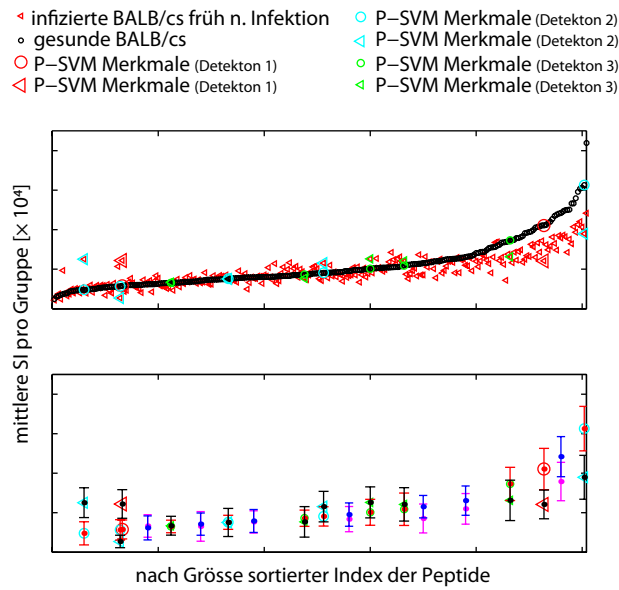
Erkenntnis, dass das Zusammenspiel der gefundenen Merkmale bei der Klassifizierung oder Kreuzvalidierungs-Vorhersage von verschiedenen Gruppen eine erhebliche Rolle spielt. Das einzelne Merkmal aus der systematischen Bibliothek trägt nur einen Bruchteil zur Differenzierung bei, erst das Zusammenspiel der einzelnen Merkmale einer Merkmalsgruppe (Detekton) macht diese zum guten Klassifikator. Das Zusammenspiel der Merkmalseigenschaften (SI der Peptide eines Detektons) resultierte z. B. in einer Normierung der hohen SI durch eine tiefere SI, wie wir in Abbildung 6.2 vorweggenommen hatten.

### 6.5.2 Detektone sind überall

Die Robustheit des Vorkommens der Detektone kann anhand ihrer Redundanz untersucht werden. Um den Verdacht zu zerstreuen, dass einzelne Detektone nur zufällig zu finden waren und sehr von einzelnen Peptiden abhingen, wurden die Peptide des ersten Detekton eliminiert und darauf nach neuen, klassifizierenden Peptiden gesucht (Abbildung 6.4). Das Testen Detektonbereinigter Datensätze verläuft folgendermassen: Nach Entnahme der Daten des ersten IgM-Ak-Detektons, das die »gesunden« und die »früh nach Infektion« Gruppen mit zwei Peptiden zu 97.7% vorhersagte, blieb ein Datensatz übrig, bei dem wir mittels Merkmalsselektion ein »Detekton 2« fanden. Das zweite Detekton konnte die zwei Gruppen mit fünf Peptiden zu 100% vorhersagen. Nach Eliminierung dieser fünf Peptide blieb ein Datensatz von  $43 \times 245$  Datenpunkten ( $(15 + 28) \text{ Samples} \times (252 - 2 - 5) \text{ Merkmale}$ ), in dem sich »Detekton 3« finden liess, das mit fünf Merkmalen die zwei Gruppen abermals zu 100% vorhersagte.

Die Varianzen der IgM- und IgG-Ak-SI aller Peptide des ganzen Datensatzes und die Varianzen innerhalb der einzelnen Gruppen werden untersucht: Die Streuung der Peptide steigt mit zunehmender SI: Mittelwert  $\mu$  und die Varianz  $\sigma$  korrelieren sehr gut (Abbildungen 6.5(a) und 6.6(a)). Die Mittelwerte der Varianzen  $\sigma$  der untersuchten Gruppen unterscheiden sich stark; mit zunehmender Zeit nach Infektion steigt der Mittelwert der Varianz innerhalb der Gruppen. Die Verschiebung des Varianzbalkens der »spät«-Gruppe (blau) in Abbildung 6.5(b) steht für eine Zunahme der Varianz in dieser Gruppe. Eine derart klare Zunahme kann nur biologischen Ursprungs sein; selbst die Varianzen der Varianzen überschneiden sich nur marginal. Die Streuung der einzelnen Samples innerhalb der Gruppen bestätigt den biologischen Ursprung (Abbildung 6.7); das Ak-Repertoire scheint mit zunehmender Infektion diverser zu werden. Eigenartigerweise ist dieser Effekt bei den IgG-Ak-Signalen eher zwischen der »gesunden« Gruppe und





**Abbildung 6.4:** Die ersten drei Detektone zur Vorhersage der »gesunden« gegenüber der »früh nach Infektion« Gruppen: Nach der Extraktion des ersten Detektors (rot) aus dem Datensatz fand sich ein zweites Detekton (türkis) und nach dessen Entnahme ein drittes Detekton (hellgrün), von denen jedes einzeln zur vollständigen Kreuzvalidierungs-Vorhersage der beiden Gruppen »gesund« vs. »früh nach Infektion« ausreichte. Die untere Darstellung zeigt Mittelwert und Varianz der Peptide der Detektone. In Dunkelblau und Violett sind zudem Mittelwerte und Varianzen von zufällig gewählten Peptiden angezeigt.

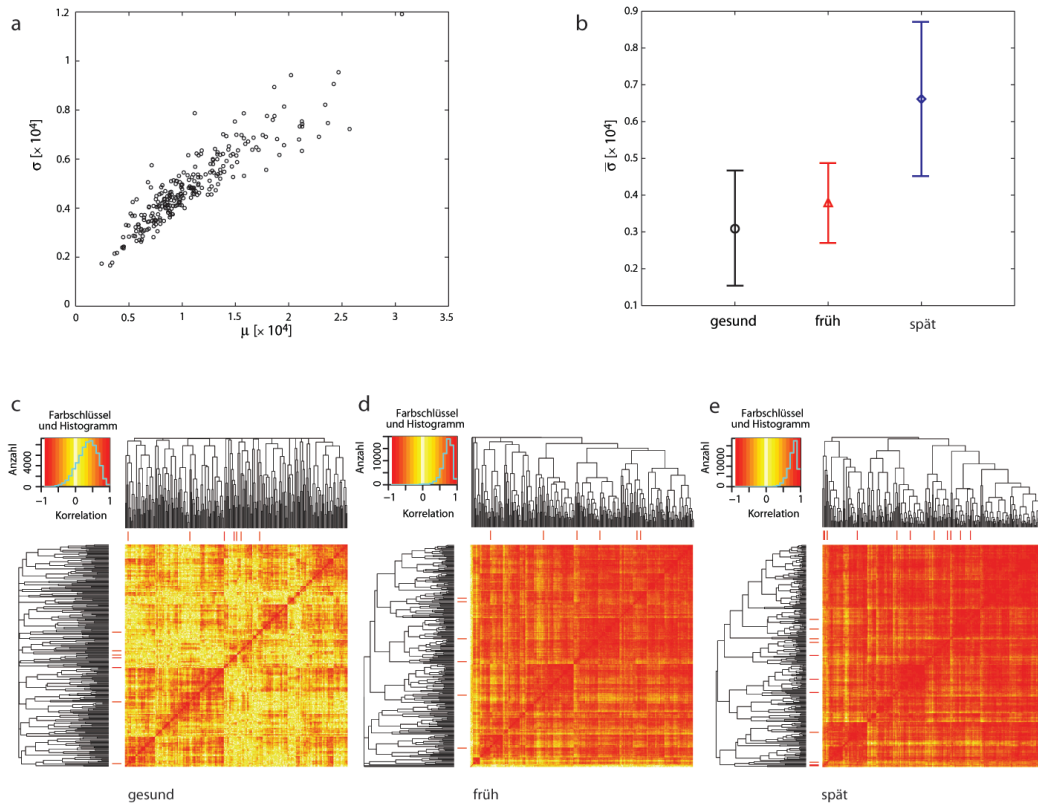
der Gruppe »früh nach Infektion« (in Abbildung 6.6(b)) sichtbar. Eine mögliche Interpretation dieser Unterschiede findet sich im Diskussionsteil 8.7.

Mit zunehmender Varianz innerhalb der Gruppen nahm auch die Korrelation innerhalb der Gruppen zu (Abbildung 6.5(c)-(e)). Die Unterschiede zwischen den SI der einzelnen Peptide nahm demnach ab. In Anbetracht identischer Farbskalen war dieser Effekt bei den IgG-Ak-SI in den Abbildungen 6.6(c)-(e) nicht ganz so drastisch. Die hellen Streifen in Abbildung 6.6(e) legen dafür nahe, dass stattdessen sehr klar voneinander abgetrennte Gruppen von Merkmalen existierten, die auch mit zunehmender Krankheit nicht miteinander korrelieren. Daraus folgt, dass sich die IgG-Ak-SI eindeutiger in unterschiedliche Gruppen aufteilen.

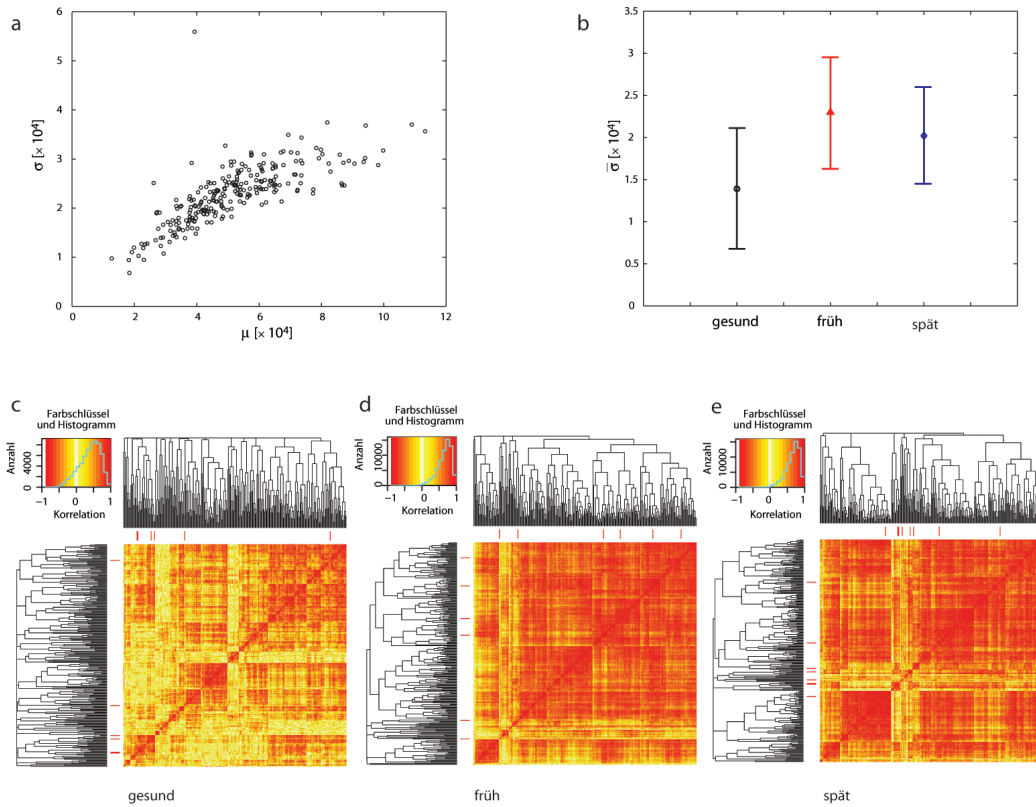
Die Unterschiede der Mittelwerte jedes Peptids zwischen den Gruppenpaaren werden anhand von Abbildungen 6.7(a)-(c) in nach SI sortierter Reihenfolge betrachtet. Auffallend ist, wie wenig sich die Mittelwerte der SI pro Peptid von einzelnen Gruppen unterscheiden. Peptide, die in einer Gruppe sehr hohe und in der andern sehr tiefe SI aufwiesen, kommen nicht vor; die Differenz der Mittelwerte der Gruppen war für alle Peptide sehr klein; Detekton-Merkmale kommen im ganzen Spektrum der SI-Höhen verteilt vor (siehe auch Abschnitt 6.5.2). Daraus folgt: Detektone kommen überall vor. Abbildung 6.7(d) zeigt des Weiteren, dass die Mittelwerte der einzelnen Samples auch innerhalb der einzelnen Gruppen sehr stark streuen, woraus wir schliessen können, dass die kleinen Differenzen der Mittelwerte nicht zur Klassifizierung ausreichten. Die SI der zwei besten Merkmale der Gruppen »gesund« und »früh« gegeneinander aufgetragen (Abbildung 6.7(e)) zeigt die hohen Korrelation der SI innerhalb der Gruppen und niedrigen Korrelationen zwischen den Gruppen. Die Untersuchung der IgG-Ak-Unterschiede in der Abbildungen 6.8 zeigt, dass die gleichen Aussagen auch auf die IgG-Ak-Signale zutreffen. Eigenschaften der SIs der zwei Peptide in Abbildung 6.7(e) wurden bereits in Abschnitt 6.1 und Abbildung 6.2 diskutiert.

## 6.6 Merkmalsselektion: P-SVM versus Wilcoxon-Vorzeichen-Rang-Test

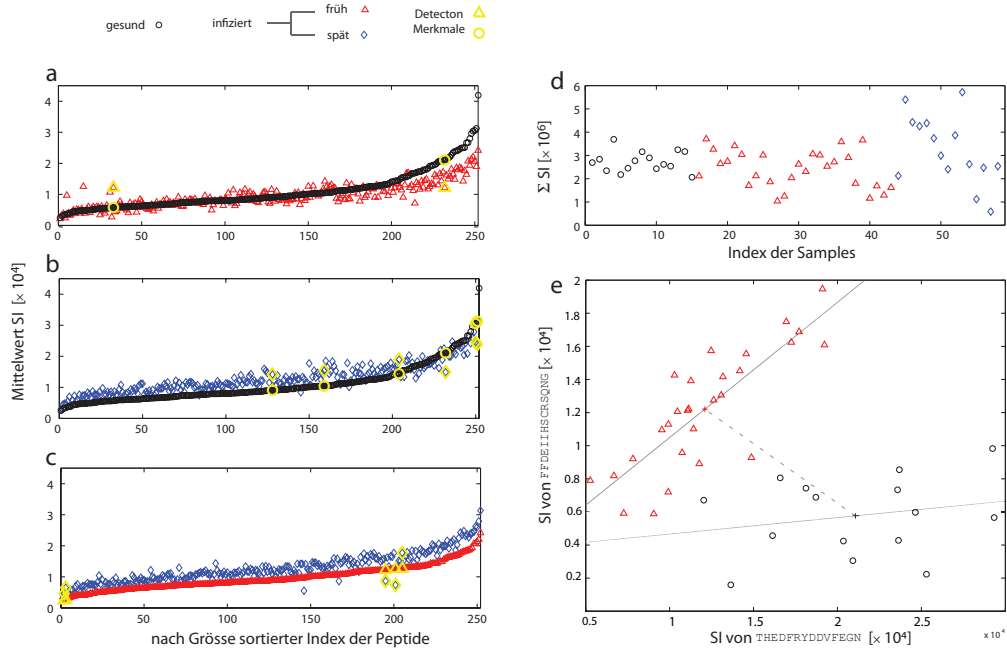
Zur Demonstration der Qualität der Merkmalsselektion mittels P-SVM folgt hier ein Vergleich der P-SVM- und der Wilcoxon-Vorzeichen-Rang-Text-Merkmalsselektion (siehe Abschnitte 3.9.6). Zur 'neutralen' Beurteilung der resultierenden Präzision der Kreuzvalidierungs-Vorhersage wurde diese mittels



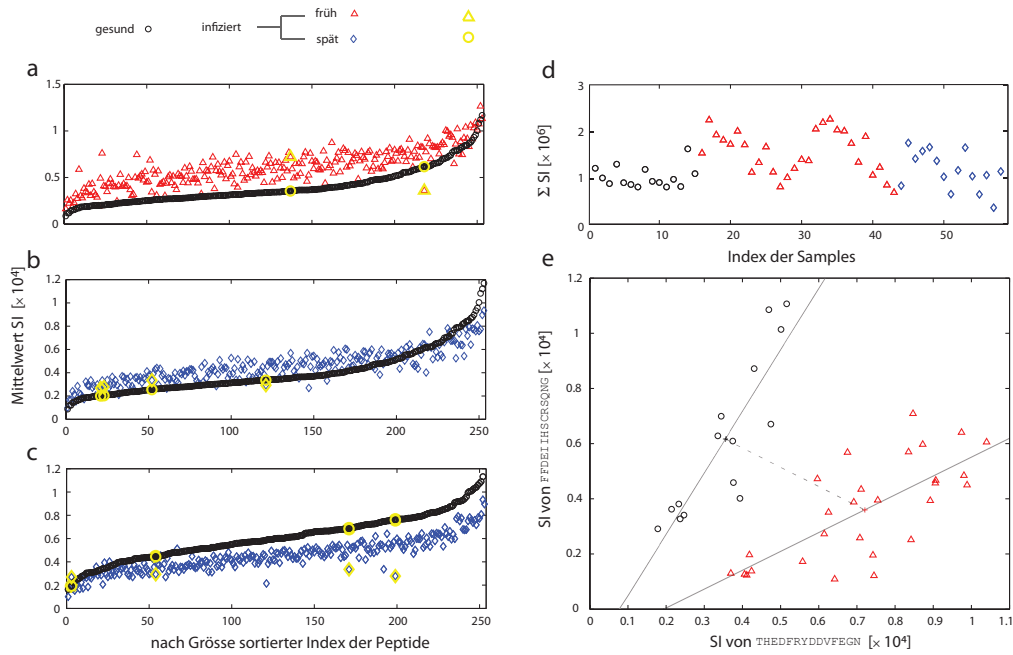
**Abbildung 6.5:** Übersicht der Eigenschaften des IgM-Ak-Datensatzes auf Glasträgern. (a) IgM-Ak-SI-Mittelwerte ( $\mu$ ) und -Varianzen ( $\sigma$ ) jedes einzelnen Peptides über alle 58 Sample-Messpunkte. (b) Mittelwert und Streuung der Varianz der IgM-Ak-SI innerhalb jeder der Gruppen. (c)-(e) Geclusterte Korrelationen der IgM-Serumantikörper-Peptid-Reaktivitäten für jede der drei Gruppen: (c) gesund, (d) früh nach Infektion und (e) spät nach Infektion.



**Abbildung 6.6:** Übersicht der Eigenschaften des IgG-Ak-Datensatzes auf Glasträgern. (a) IgG-Ak-SI-Mittelwerte ( $\mu$ ) und -Varianzen ( $\sigma$ ) jedes einzelnen Peptides über alle 58 Sample-Messpunkte. (b) Mittelwert und Streuung der Varianz der IgG-Ak-SI innerhalb jeder der Gruppen. (c)-(e) Geclusterte Korrelationen der IgG-Serumantikörper-Peptid-Reaktivitäten für jede der drei Gruppen: (c) gesund, (d) früh nach Infektion und (e) spät nach Infektion.



**Abbildung 6.7:** Eigenschaften der Detekton-Merkmale der IgM-Antikörperbindungen. (a)-(c) Die mittleren IgM-Ak-SI der Detekton-Merkmale (gelb) im Vergleich zu allen anderen mittleren SI der Peptide pro Gruppe. Die Mittelwerte der SI von jeweils zwei Gruppen wurden in aufsteigender Reihenfolge der Mittelwerte innerhalb einer Gruppe sortiert und die Positionen der besten Detekton-Merkmale aus Tabelle 6.2 markiert. (c) suggeriert, dass die Differenzen der Mittelwerte zwischen den Gruppen zur Klassifizierung ausreichen könnten. (d) zeigt jedoch, dass die Streuung der Mittelwerte viel zu gross war; viel grösser als die Differenzen zwischen den Gruppen. (e) zeigt die einzelnen IgM-Ak-SI der zwei Detekton-Merkmale aller gesunden und früh-nach-Infektion Seren. Die Streuung zeigt, dass die Korrelation innerhalb der Gruppen sehr hoch war, dafür aber sehr klein zwischen den Gruppen. Die Samplegruppen sind: gesunde ( $\circ$ ) BALB/c-Mäuse und infizierte BALB/c-Mäuse früh ( $\triangle$ ) und spät ( $\diamond$ ) nach Infektion.



**Abbildung 6.8:** Eigenschaften der Detekton-Merkmale der IgG-Antikörperbindungen. (a)-(c) Die mittleren IgG-Ak-SI der Detekton-Merkmale (gelb) im Vergleich zu allen anderen mittleren SI der Peptide pro Gruppe. Die Mittelwerte der SI von jeweils zwei Gruppen wurden in aufsteigender Reihenfolge der Mittelwerte innerhalb einer Gruppe sortiert und die Positionen der besten Detekton-Merkmale aus Tabelle 6.2 markiert. (a) und (c) suggerieren, dass die Differenzen der Mittelwerte zwischen den Gruppen zur Klassifizierung ausreichen könnten. (d) zeigt jedoch, dass die Streuung der Mittelwerte viel zu gross war; viel grösser als die Differenzen zwischen den Gruppen. Die Streuung der Mittelwerte der früh-nach-Infektion-Seren ist am höchsten hier. (e) zeigt die einzelnen IgG-SI der zwei Detekton-Merkmalen aller gesunden und früh-nach-Infektion-Seren. Die Streuung zeigt, dass die Korrelation innerhalb der Gruppen sehr hoch war, dafür sehr klein zwischen den Gruppen. Die Samplegruppen sind: gesunde (o) BALB/c-Mäuse und infizierte BALB/c-Mäuse früh ( $\Delta$ ) und spät ( $\diamond$ ) nach Infektion.

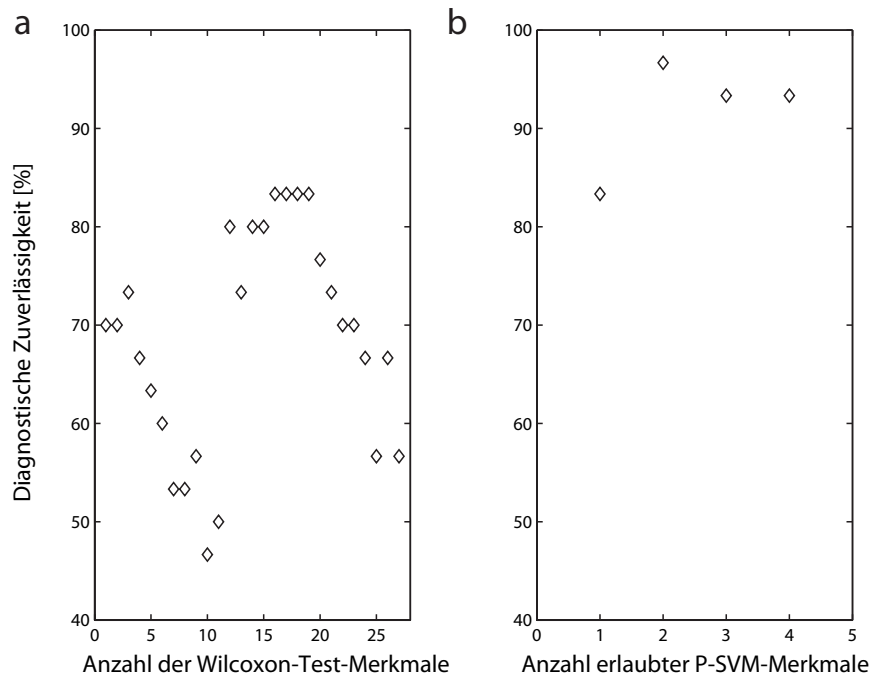
LDA auf den selektierten Merkmalen durchgeführt. Abbildung 6.9 vergleicht die Qualität der Merkmalsselektion von P-SVM und Wilcoxon-Vorzeichen-Rang-Test (kurz: »Wilcoxon-Test«). Bei den untersuchten Gruppen handelte es sich um die gesunden und die infizierten »spät nach Infektion«. Der Wilcoxon-Test ordnet die Peptide einzeln nach der Wahrscheinlichkeit, dass sie die zwei untersuchten Gruppen zu unterscheiden vermögen. Im vorliegenden Fall, in dem kein einzelnes Peptid 100% zu unterscheiden vermochte, zeigte sich, dass diese Sortiermethode nicht in der Lage war, eine Gruppe gut vorhersagender Peptiden zu finden. Die LDA-Kreuzvalidierung sagt jedes der 30 Seren der zwei untersuchten Gruppen aufgrund des Modells, das die andern 29 mit der entsprechenden Zahl Peptide (Merkmale) gebildet hatten, vorher. Die Fehlerrate (pro Merkmalszahl) wird jeweils mittels der 30 Vorhersagen berechnet. Die Zahl der Peptide ist in Abbildung 6.9 gegen die Präzision der LDA-Kreuzvalidierung aufgetragen. Die Zuverlässigkeit der LDA-Vorhersage mittels der von P-SVM gewählten Peptide ist viel höher. Tabelle 6.3 zeigt allerdings, dass die bestrangierten Peptide von beiden Methoden als wichtig selektiert wurden.

P-SVM Reihenfolge	Wilcoxon-Test Reihenfolge	Peptidsequenz
1	3	PDWLTILYSFWFES
2	17	LLMKYAFDDRQAEW
3	2	QSFTQKNNAIQFNK
4	12	THEDFRYDDVFEGN
5	7	LDEEVFTFENLRIY

**Tabelle 6.3:** Peptide, die von P-SVM selektiert wurden für die höchsten diagnostischen Zuverlässigkeiten zwischen gesunden BALB/c-Mäusen und *H. polygyrus*-infizierten spät nach Infektion. Peptide, die von P-SVM ausgewählt wurden, haben auch einen hohen Rang im Wilcoxon-Test.

## 6.7 Multiklassen-Klassifizierung in zwei Dimensionen

Die zeitliche Abfolge der Infektion kann anhand der Immunreaktionen sowohl im IgM-Ak- als auch im IgG-Ak-Repertoire gezeigt werden. Abbildung 6.10 zeigt zweidimensionale LDA-Multiklassen-Klassifizierungen mit Peptiden, die mittels P-SVM-Kreuzvalidierungs-Merkmalsselektion zwischen allen Gruppenpaaren aus Tabelle 6.1 gewählt wurden. Die zeitliche Abfolge der In-

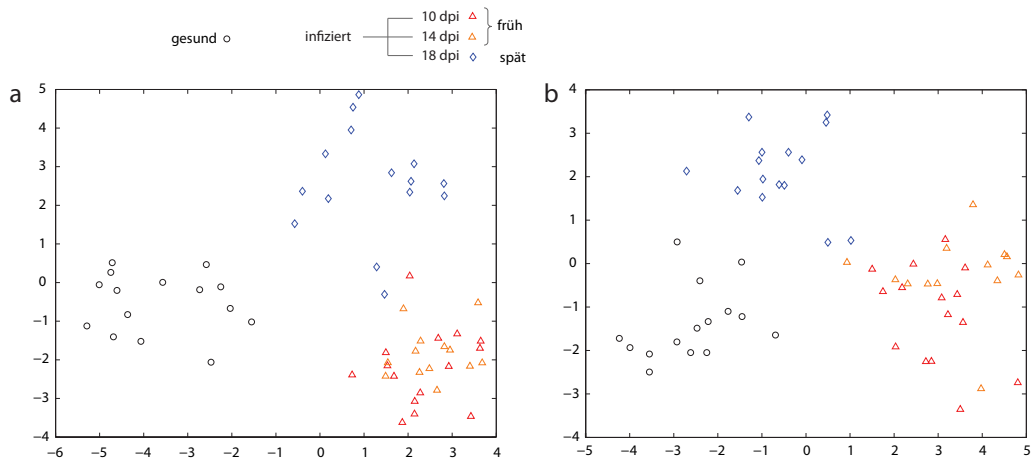


**Abbildung 6.9:** Vergleich der Merkmalsselektion-Methoden Wilcoxon-Test und P-SVM. (a) Wilcoxon-Vorzeichen-Rang-Test- und (b) P-SVM-Merkmalsselektion zwischen gesunden und infizierten BALB/c-Mäusen zu einem späten Zeitpunkt. Die Sicherheit der Vorhersage wurde mittels vollständiger Kreuzvalidierung per LDA abgeschätzt. Die Angabe der Anzahl auf der  $x$ -Achse steht in (a) für die Anzahl der Peptide, die aus der Wilcoxon-Test-Rangfolge benutzt wurden und in (b) für die Anzahl der Peptide, die P-SVM maximal selektieren durfte.



fektion ist durch die Auftrennung der einzelnen Gruppen gut ersichtlich. Die diagnostische Zuverlässigkeit (mittels LDA auf Label-Austausch-Datensätzen ermittelt) liegt bei 93.1% für die IgM-Signale mit 11 Peptiden und bei 94.8% für die IgG-Signale mit 10 Peptiden.

Die LDA-Multiklassen-Klassifizierung ist hier mit viel deutlicherer Auftrennung, ohne Datennormierung und mit der Hälfte der Peptidzahl, die für die Zellulosemembrandaten benötigt wurde (Abschnitt 5.4) möglich (zugegebenermaßen mit einer Gruppenzahl weniger). Die Resultate der Glasträger-Serumantikörper-Reaktivitäts-Profil-Unterscheidungen übertreffen die der Zellulose bei Weitem.



**Abbildung 6.10:** LDA-Multiklassen-Klassifizierung (a) mit den IgM-Ak-SI von elf Peptiden und (b) mit den IgG-Ak-SI von zwölf Peptiden. Die Sample-Gruppen sind: gesunde (o) BALB/c-Mäuse und infizierte BALB/c-Mäuse früh ( $\Delta$ ) und spät ( $\diamond$ ) nach Infektion.



# Kapitel 7

## Klassifizierung humaner Serumantikörper

Das Kapitel behandelt die Klassifikation des extern akquirierten Pertussis-Datensatzes der humanen Serensample-Inkubationen. Der Datensatz ist in Abschnitt 3.8 näher beschrieben. Der Datensatz ist zweiteilig und das Ziel der Untersuchungen war die Klassifizierung der beiden Gruppen des Datensatzes. Die Untersuchung humaner IgG-Antikörper-Reaktivitäts-Profilen auf Peptidbibliotheken auf Glasträgern bietet eine interessante, anwendungsnahe Ergänzung zu den Resultatsanalysen der murinen Antikörper-Reaktivitäts-Signale aus Kapitel 6.

### 7.1 Klassifizierung von humanen Seren von Pertussis-geimpften und Placebo-behandelten Individuen

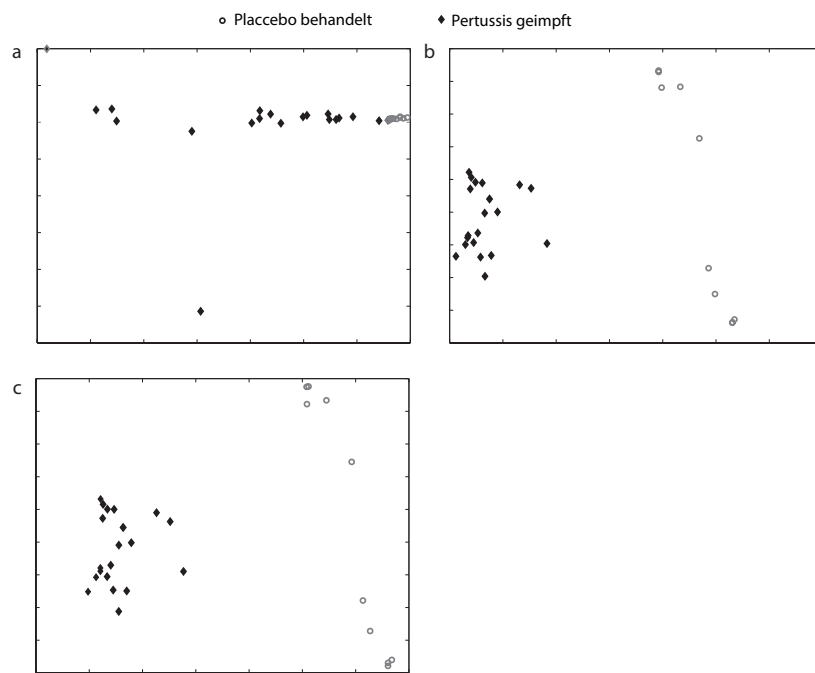
Der hier untersuchte, extern akquirierte Datensatz besteht aus vorverarbeiteten Signalintensitäts-Vektoren der Sereninkubationen von 20 Pertussis-geimpften und 10 Placebo-behandelten Probandensamples. Pro Probandensample standen 2934 Signalintensitäten zur Verfügung, die von den Bindungsmessungen der Serumantikörper an Peptidspots stammten. Der Datensatz war bereits aufbereitet, d.h. Kontrollsignale und falsch-positive Signale, die von den sekundären Antikörpern rührten, waren eliminiert. Auf der Suche nach der idealen Normierung, welche die Verteilung der Daten optimiert, haben wir die Daten unterschiedlich normiert und dann die ersten beiden Hauptkomponenten der PCA der jeweiligen Normierung betrachtet. Abbildung 7.1 zeigt die unterschiedlich normierten SI der 30 getesteten Seren

nach der Hauptkomponentenanalyse (PCA): Der Datensatz separierte sich unabhängig von der Normierung bereits mit PCA. Abbildung 7.1(a) zeigt die ersten beiden Hauptkomponenten der PCA der nicht normierten Daten, (b) zeigt die Hauptkomponenten mittels Median-Subtraktion und  $\sigma$ -Division normierten Daten (nach Formel 3.4) und (c) die Hauptkomponenten der auf 1 normierten Daten nach Formel 3.5. Alle Normierungen klassifizieren die Daten vollständig in der ersten Hauptkomponente, d.h. allein aufgrund der grössten Varianzen. Die Anteile der Hauptkomponenten an der Varianz sind in der ersten Komponente 46.5 % (a), 28.6 % (b) und 28.4 % (c) und in der zweiten Komponente 13.0 % (a), 8.7 % (b) und 8.4 % (c). Die Daten klassifizieren sich in der ersten Hauptkomponente, unabhängig von der Normierung. Während sich die zwei Gruppen der Daten im nicht normierten Fall in Abbildung 7.1(a) räumlich kaum auftrennen (alle Placebo-Datensätze befinden sich oben rechts auf einem Haufen), bieten die normierten Datensätze (b) und (c) viel deutlichere Auftrennungen.

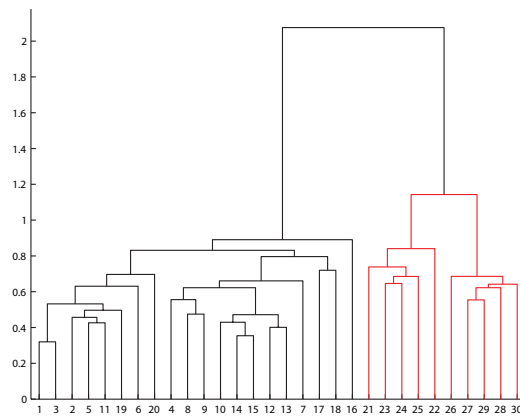
P-SVM Merkmalsselektion und Vorhersage konnte aufgrund der Datenmenge nur auf einer reduzierten Zahl der Merkmale (Peptiden) durchgeführt werden. Die Auswahl von 300 zufällig gewählten Proben sagte die 30 Datensamples mit vier dieser Proben (Merkmale) mittels leave-one-out-Kreuzvalidierung mit 93.2%iger Präzision und einer Signifikanz von  $p < 0.01$  vorher (Signifikanzberechnung in Abschnitt 3.9.10 beschrieben). Mit Merkmalsselektion gelingt also eine sehr gute Vorhersage der zwei Gruppen des normierten Pertussis-Datensatzes.

Des Weiteren zeigt Abbildung 7.2, dass das Aufgruppieren (d.h. Clustern) der vorverarbeiteten, normierten Datensätze der geimpften und der Placebo-behandelten Gruppe mit dem »Ward«-Algorithmus (Abschnitt 3.9.7) fehlerfrei gelingt. Fehlerfreies Clustern mit einer Standard-Clustering-Methode ist eine Eigenschaft, die nur sehr gut separierbare Datensätze aufweisen.

Die gute Klassifizierbarkeit des Datensatzes liefert einen Beweis, dass sich unsere Methoden auch zum Auftrennen humaner Serumantikörper-Reaktivitäts-Profile eignet. Die Varianz der humanen Daten erlaubt die gleiche Vorgehensweise der Diagnostik, die für die murinen Daten in Kapitel 6 gezeigt wurde. Eine Diskussion der Problematik dieses Datensatzes folgt in Abschnitt 8.6.



**Abbildung 7.1:** Hauptkomponentenanalyse der Signalintensitäten von 20 Pertussis-geimpften und 10 Placebo-behandelten Seren. (a) ohne Normierung, die Placebo behandelten befinden sich in der rechten oberen Ecke und sind kaum zu erkennen, nur ein einzelner Geimpfter macht die zweite Dimension aus. (b) normiert nach Formel 3.4 und (c) nach Formel 3.5. Die Beschriftung der Größenordnung der Achsen ist weggelassen; das Ziel dieser Abbildungen ist eine Veranschaulichung der Verteilungen der Serensamples in der Ebene aufgrund der Normierung.



**Abbildung 7.2:** Clusteranalyse mittels Ward-Algorithmus der SI von 20 Pertussis-geimpften und 10 Placebo-behandelten Seren, normiert nach Formel 3.4.

# Kapitel 8

## Schlussfolgerungen und Ausblicke

Die Arbeit weist darauf hin, dass sich eine nicht-invasive Diagnostik aufgrund von Reaktivitäts-Profilen von ganzen Serumantikörper-Repertoires zu einer robusten und verlässlichen Anwendung für verschiedenste Krankheitsbilder entwickeln kann. Darüber hinaus wirft sie Fragen bezüglich des Sinns der Suche nach »Biomarkern« und ihrer hohen Bindungskonstanten auf.

### 8.1 Erreichte Ziele

Die verwendeten Technologien und Verfahren sind sowohl auf der experimentellen wie auch auf der Ebene der Datenauswertung soweit ausgereift, dass ein nicht-invasives Diagnostikverfahren mit Anwendungsgebieten, die unabhängig von bekannten Antigenen sind, in erreichbare Nähe rückt. Ein wichtiges Resultat ist die Tatsache, dass wir für die Differenzen zwischen allen untersuchten Gruppen von verschiedenen Seren (die unterscheidbar sind) kleine Detektone (Abschnitt 6.5) finden, die sich zur Klassifizierung und Vorhersage eignen: Die meisten Serengruppen konnten mit einem erstaunlich kleinen Detekton mit grosser Zuverlässigkeit und hoher Signifikanz klassifiziert werden.

Zusammenfassend sind die grössten Errungenschaften dieser Arbeit die sorgfältige Datenauswertung und -analyse, vor allem die Merkmalsselektion, die breite Anwendbarkeit der unvoreingenommenen Probekbibliotheken aufgrund ihrer Unabhängigkeit von der Fragestellung sowie die Verwendung von synthetisch hergestellten Peptiden. Die Tatsache, dass sowohl die biologischen Experimente als auch die statistischen Analysen in derselben Gruppe und von

zumindest teilweise sogar den gleichen Personen bearbeitet wurden, ist eher aussergewöhnlich und hat wohl einen nicht zu unterschätzenden Beitrag zum Gelingen der Arbeit geleistet.

Die Tatsache, dass die Klassifizierung und Kreuzvalidierungs-Vorhersagen, wenn überhaupt, dann nicht nur mit den zufällig generierten Sequenzen, sondern darüber hinaus mit erstaunlich kleinen Anzahlen von Peptiden realisierbar sind, wirft Fragen bezüglich der Zusammenstellung der Probenbibliothek auf. Die Probenbibliotheks-Grössen von 156 (Zellulosemembranen) bzw. 255 (Glasträger) unterschiedlichen Peptidsequenzen erscheint in Anbetracht der Grösse des potenziellen Peptid-Interaktions-Raumes als sehr klein. 255 Peptide stellen allerdings mehr als 255 Bindungsstellen dar: Die Zahl der Aminosäuren, die an einer Bindungsstelle von Antikörpern und Antigenen beteiligt ist, variiert in Menge und Abstand voneinander. Bekannte Interaktions-Längen sind kürzer als 14 Aminosäuren, weswegen 255 14-mal mehr ein Mehrfaches von 255 potenziellen Antikörper-Bindungsstellen offerieren. 255 unterschiedliche Peptidsequenzen scheinen zumindest für die Klassifizierung des murinen Datensatzes in Kapitel 6 eine genügend grosse Menge zu sein, um eine Unterscheidung aufgrund von IgM- und IgG-Antikörper-SI zu ermöglichen. IgM Antikörper sind des Weiteren als äusserst polyreaktiv beschrieben, und zur Unterscheidung der verschiedenen analysierten Gruppen muss anscheinend nicht das ganze Spektrum der Reaktivitäten ausgelotet werden.

## 8.2 Zusammensetzung der Probenbibliotheken

Die Frage nach der optimalen Grösse und der optimalen Diversität der Bindungspartner der Probenbibliothek, welche die Unterschiede der untersuchten Serengruppen mit minimalstem Aufwand zu finden erlauben, ist zurzeit nicht allgemein beantwortbar. Wir können allerdings davon ausgehen, dass eine solche ideale Bibliothek nicht ausschliesslich aus zufälligen Sequenzen bestehen muss. Eine ideale, breit anwendbare Bibliothek böte eine maximale Diversität, die den ganzen (nicht hochaffinen) Bindungsraum mit möglichst kleiner Redundanz abdecken würde (um Samples und Sample-Material einzusparen). Die Vergrösserung der Redundanz böte andererseits den Vorteil, dass ähnliche oder gleiche Bindungspartner robustere Signale lieferten. Die Grösse des chemischen Raumes und den Anteil, den das Immunsystem davon abdeckt, zu bestimmen, übersteigt das in dieser Arbeit Machbare jedoch.



Des Weiteren muss eine solche ideale Bibliothek gar nicht zwingend aus Aminosäurenketten bestehen. Bibliotheken anderer chemischer Komplexe, wie z. B. kleine organische Moleküle oder Zuckerringe [83], erfüllen die Unterscheidungsaufgaben offensichtlich auch. Die pharmazeutische Industrie zieht auf der Suche nach pharmakologisch wirksamen Bindungspartnern die kleinen chemischen Moleküle den natürlichen vorkommenden vor. Kleine chemische Moleküle haben gegenüber Ketten von natürlichen Aminosäuren den Vorteil, dass sie schlechter vom Metabolismus umgesetzt und abgebaut werden. Ihr geringeres Molekulargewicht ermöglicht zudem, dass die kleinen Moleküle zumindest teilweise viel einfacher in der Lage sind, biologische Membranen zu durchqueren, eine wichtige Voraussetzung für oral verabreichbare Medikamente. Der Vollständigkeit halber soll an dieser Stelle angemerkt werden, dass viele der kleinen organischen Moleküle auch Nachteile wie z. B. schlechte Wasserlöslichkeit aufweisen können. Solche Vor- und Nachteilabwägungen von Probenbibliotheken spielen aber nur eine Rolle, wenn das Ziel der Untersuchung eine »Hit-Suche« ist, wie z. B. die Suche nach Biomarkern oder Medikamenten. Derartige Suchen werden sinnvollerweise aber nicht mit Probenbibliotheken mit systematischer Zusammensetzung (siehe Abschnitt 1.5.2) vorgenommen, denn das Auffinden von hochaffinen Bindungspartnern wird unabhängig von der Weiterentwicklung unserer Methode auf ein detailliertes Vorwissen der molekularen Zusammenhänge angewiesen sein.

## 8.3 Normierung und Nicht-Normierung

Die Argumentation des Verzichtes auf eine Normierung, weil die Daten mit keiner der untersuchten Normierungen besser klassifizierbar werden, ist anfechtbar (Abschnitt 6.1). Eine Normierung sollte selbstverständlich objektiv sein. Eine sinnvolle Normierung würde auf Inkubationswiederholungen basieren. Die vorliegenden Wiederholungen in Abbildung 4.9 haben höhere Reproduzierbarkeiten als die analysierten Daten, weil sie auf nur drei Trägern am selben Tag unter idealen Bedingungen gemessen wurden. Idealerweise würden Repetitionen parallel zu den Experimenten gemessen. Die Zahl der parallel durchführbaren Experimente ist zurzeit beschränkt und erlaubt keinen solchen Ansatz (der Glasträgerdatensatz in Kapitel 6 wurde an zwei Tagen gemessen, was nur dank der Unterteilung der Glasträger in fünf Subarrays möglich war; eine Aufdrucktechnologie, die inzwischen leider nicht mehr erhältlich ist).

Den Standpunkt »lieber keine Normierung als eine, die den Informationsgehalt der Daten derart reduziert, dass die gesuchten biologischen Unterschiede schlechter gefunden werden«, können wir aber mit gutem Gewissen vertreten. Eine optimale Normierung für Glasträgerinkubationen zu finden ist ein wichtiges Ziel weiterer Arbeiten.

## 8.4 Diagnose und Infektionsverlauf

Leave-one-out-Kreuzvalidierungs-Vorhersagen als Diagnose zu bezeichnen ist ebenfalls anfechtbar. Idealerweise müssten wir neue Messungen mit neuen Serensample-Sätzen und neuen Trägern herstellen und diese dann aufgrund der Modelle vorhersagen, die vorher aufgestellt worden waren. Derartige Experimente und Auswertungen werden zurzeit vorgenommen; Technologie und Lieferbarkeit der Glasträger mit möglichst identischen Merkmalen bieten noch immer viele Schwierigkeiten. Der P-SVM-Algorithmus zur Merkmalsselektion wird aber, wenn richtig angewandt und interpretiert, aufgrund seiner Robustheit auch für hoch variierende Datensätze Merkmale finden, die sich zur Unterscheidung eignen (vorausgesetzt messbare biologische Unterschiede existieren). Ich gehe davon aus, dass die Datensätze der *H.-polygyrus*-infizierten Mäuse von 10 dpi und 14 dpi zu wenig Unterschiede in ihren Ak-Repertoires enthielten, um mit den hier verwendeten Technologien biologische Effekte messen zu können. Der Verlauf der Infektion war zu den Zeitpunkten der frühen Seren-Entnahmen im Antikörperrepertoire noch nicht ausreichend sichtbar.

Generell unterscheiden sich die Resultate der Antikörper-Reaktivitäts-Profilen der einzelnen untersuchten Gruppen mit unerwarteter Klarheit und Präzision (vor allem die der Glasträger-Bindungsmessungen). Der Versuch aufgrund kleiner Zahlen von tief-affinen Bindungen Unterschiede zu finden ist über alle Erwartungen gut verlaufen. Die detaillierteren Eigenschaften der Peptidarrays auf Glasträgern und der Detektone zur Unterscheidung verschiedener Immunitätszustände weiter zu untersuchen, wird eine hochinteressante zukünftige Arbeit.

## 8.5 Vergleich der Inkubationsträger

### Zellulosemembranen und Glasträger – Makro versus Mikro

Die in dieser Arbeit benutzten Trägertechnologien zur Messung verschiedener Serensamples gegenüber Peptidbibliotheken basieren auf der SPOT Synthese, einer synthetischen Synthese von Peptidbibliotheken auf Zellulosemembranen. Die Bindungsmessung (d.h. Inkubation) der Serensamples gegenüber der Peptidbibliotheken wird direkt auf den Makroarrays (Zellulosemembranen) oder auf Mikroarrays (Peptide werden ausgelöst und übertragen) durchgeführt.

Die Technologie der Seruminkubation auf Zellulosemembranen ist technisch weniger anspruchsvoll, einfacher in der Handhabung, aber weniger reproduzierbar als die Inkubation auf Glasträgern. Zur Herstellung der Glasträger werden die Peptide ebenfalls per SPOT Synthese [86], [17] hergestellt, allerdings auf Spots mit grösseren Durchmessern, was zu grösseren Peptidmengen und reineren Synthesen führt. Peptide, die zum Auftragen auf Glasträger bestimmt sind, werden nach der Synthese aus der Zellulose ausgelöst und anschliessend mittels Kontakt- oder Nichtkontakt-Plottern punktweise auf Glasträger aufgetragen. Die Glasträgertechnologie erlaubt eine 100- bis 1000fach höhere Dichte verschiedener Spots (d.h. Peptide) pro Fläche und ein deutlich kleineres Inkubationsvolumen (siehe Tabelle 1.1). Kleinere Inkubationsvolumen sind vor allem für Seren muriner Abstammung, die meist nur in kleinen Mengen verfügbar sind, sehr vorteilhaft. Die Herstellung von Mikroarrays auf Glasträgern ist aber technisch derart anspruchsvoll, dass die peptidbeladenen Glasträger nur von spezialisierten, erfahrenen Laboren hergestellt werden können. Die Abhängigkeit von einer externen Firma zur Herstellung der peptidbeladenen Glasträgern erwies sich, trotz guter Zusammenarbeit, als ein aufwendiger, zeitlich und technologisch schwer berechenbarer Faktor.

Weitere Unterschiede zwischen den Technologien der Inkubation auf Zellulosemembranen und Glasträgern betreffen die Markierung der sekundären Antikörper: Während die Bindungen der Antikörper an Peptide auf der Zellulosemembran mit sekundären Antikörpern, die mit Peroxidase markiert sind, nachgewiesen und ausgelesen werden, tragen die sekundären Antikörper der Glasträger Fluoreszenzmoleküle, die mit entsprechenden Lasern angeregt und ausgelesen werden. Die Fluoreszenzmarkierung hat den Vorteil, dass sie unabhängig von biologisch aktiven Molekülen, wie Enzyme, funktioniert

und daher reproduzierbarer und aufgrund der kleineren Grösse der Fluorophore auch genauer ist. Darüber hinaus besitzt der Laserscanner zwei (oder mehr) Laser mit unterschiedlichen Wellenlängen zur Anregung der Fluorophoren (inkl. den entsprechenden Filtern), wodurch die Glasträger mit unterschiedlichen sekundären Antikörpern *gleichzeitig* inkubiert werden können. Das Peroxidase-Signal reicht nur zur Bindungsdetektion einer Klasse. Das parallele Auslesen verschiedener Sekundärantikörper erlaubt die Analyse der Bindung verschiedener Ig-Klassen in einem Versuch, wodurch Material, Arbeit und Zeit eingespart werden. Die Glasträgertechnologie hat aufgrund all dieser Faktoren zumindest für Hochdurchsatzanalysen eine deutlich vielversprechendere Zukunft.

### **8.5.1 Vergleich verschiedener Methoden der Herstellung von Zellulosemembranen**

Der Unterschied der Signal-Rausch-Verhältnisse zwischen der manuellen und der vollautomatischen Synthese der Zellulosemembranen (die in Abschnitt 3.3.2 beschrieben sind) ist grösser als erwartet. Die manuelle Herstellung eines 15mer-Peptidarrays auf einer Zellulosemembran von maximal möglicher Grösse (siehe Abschnitt 3.3.1) ist ein zeitaufwendiges Verfahren, welches eine Person über ca. sieben Arbeitstage beschäftigt. Das Auftragen einer Pipettierreihe (d.h. das Aufpipettieren aller den Sequenzvorgaben entsprechenden Aminosäuren einer Position) ist ein Vorgang von mehreren Stunden. Die Peptide können daher pro Arbeitstag nur um zwei bis drei Aminosäurekopplungen verlängert werden können. Die manuelle Synthese ergibt jedoch reinere Peptide. Die (uns bekannten) Anlagen zur vollautomatischen Synthese sind darüberhinaus fehleranfällige Maschinen. Doch trotz der technischen Anfälligkeiten verläuft eine vollautomatische Synthese in der Regel viel schneller, vor allem weil der Roboter 24 Stunden pro Tag arbeitet.

## **8.6 Extern akquirierte Datensätze**

Bei den Resultaten in Kapitel 7 handelt es sich um bisher unpublizierte Daten, zu denen wir nur unvollständige Informationen bezüglich Probanden, Serien, Inkubationen und Akquisen der Signalintensitäten besitzen. Abschnitt 1.3.2 beschreibt die Pertussis-Infektion und Abschnitt 3.8 lieferte eine Übersicht des Datensatzes, den wir zur Analyse erhalten haben. Unbekannte Daten auszuwerten, erschwert die Analyse und Interpretation der Ergebnisse erheblich. Die Schlussfolgerungen der entsprechenden Resultate werden

trotzdem hier nochmals besprochen:

Die Resultate der Pertussis-Daten-Klassifikationen sind sehr gut. Je nach Gewichtung der hier folgenden Aussage kann man aber deren Vertrauenswürdigkeit stark anzweifeln: Laut Protokoll sind die Inkubationen aller Placebo-Seren am gleichen Tag gemacht worden und erst 10 Tage später erfolgten die Inkubationen von allen »DTPa2«- und »DTPa5«-geimpften. Dazu halten wir fest, dass wir

1. nicht wissen, was »Vac status DTPa2« und »DTPa5«, abgesehen von »Pertussis-geimpft«, bedeutet, und
2. nicht ausschliessen können, dass die gute Klassifizierbarkeit zumindest teilweise auf die unterschiedlichen Inkubations-Tage der zwei Samplegruppen (Placebo-behandelt vs. Pertussis-geimpft) zurückzuführen ist. Die Klassifikation von Inkubationen der verschiedenen Gruppen an zwei verschiedenen Tagen unterscheidet im schlimmsten Fall nur Artefakte.

Abgesehen von diesen Unstimmigkeiten zeigt der Pertussis-Datensatz, dass die Reaktivitäts-Profile von Seren gesunder Menschen nicht stärker untereinander variieren als die der Mäuseseren. Wir können die Vorgehensweise der Datenanalyse der Mäuseseren aus Kapitel 6 und die gewonnen Erkenntnisse bezüglich der Mäuseseren-Unterscheidungen voraussichtlich direkt auf humane Seren übertragen.

## 8.7 Protektone und Detektone

Die Protekton-Theorie wurde oft zitiert (in 62 Artikeln und Übersichtsartikeln) und kommentiert [15], [50]. Der Theorie der Aufsummierung einer gleichen minimalen Schutzeinheit könnte man entgegenhalten, dass nicht alle Arten von Lebewesen mit der gleich grossen Anzahl von potenziellen Eindringlingen konfrontiert werden. Deshalb ist die Annahme, dass die Immunsysteme verschiedener Säugetiergattungen im Grossen und Ganzen nicht in ihrer Komplexität variieren, eine sehr vereinfachte Annahme. Des Weiteren wurde gezeigt, dass die Varianzen der Immunsysteme zu verschiedenen Zeitpunkten im Verlauf des Lebens (während der Ontogenese) sehr hoch sind [44].

Ein sehr schöner Aspekt der Protekton-Theorie ist jedoch die ergänzende Einheit und das Zusammenspiel, das alle Teilnehmer gemeinsam erbringen müssen, um zusammen ein Protekton zu bilden. Die minimale Einheit und deren Zusammenspiel im Auftrag des Schutzes leitet auch zur minimalen Einheit und dem Zusammenspiel im Auftrag der Erkennung (Detektion) über.

Wir haben nur Detektone gefunden, die aus erstaunlich wenigen Peptiden bestanden.

Eine Ausweitung des Detekton-Begriffes auf die Suche nach Biomarkern mittels voreingenommener Probenbibliotheken und auf deren »Hits« würde wenig Sinn machen: Einen einzelnen Binder Detekton zu nennen ist von geringem Nutzen. Der Begriff des Detektons macht vor allem Sinn als Bezeichnung für Detektionseinheiten, die mittels unvoreingenommener Bibliotheken gefunden werden. Das Detekton besteht dann aus tief- bis moderat-affinen Bindern, die zusammenspielen, sich gegenseitig normieren und in der Detektion der Unterschiede unterstützen (Abbildung 6.2).

### 8.7.1 Weitere Untersuchungen der Detektone

Welche Konzentrationsänderungen von einzelnen Antikörpern die gefundenen Detektone beeinflussen können, ist mit den vorliegenden Daten nicht zu beantworten. Die Grössenordnung der Zahl verschiedener Klone von Antikörpern, welche die Bindungsmusterdifferenzen beeinflussen, könnte man mit monoklonalen Antikörpern und deren Mischungen extrapolieren. Ein anderer Ansatz wäre die Untersuchung der Änderungen der Bindungen eines gegebenen Serum nach Zugabe verschiedener Konzentrationen eines monoklonalen Antikörpers. Eine weiterführende Frage ist auch, ob die Zugabe der richtigen Konzentration eines einzelnen Antikörpers ausreichen würde, um eine Verschiebung der SI-Varianzen wie in Abbildung 6.5(b) zu erhalten. Oder die Frage, ob die Zugabe eines Antikörpers in verschiedenen Konzentrationen ausreichen würde, um die Korrelationen innerhalb der Gruppen zu erhöhen (wie in Abbildungen 6.5(c)-(e)).

### 8.7.2 IgM- und IgG-Detektone im Vergleich

Die Unterschiede der Varianzen der SI zwischen den verschiedenen Ig-Klassen sind ebenfalls hoch. Die SI der IgM- und IgG-Antikörper sind in den Abbildungen 6.5(b) und 6.6(b) gezeigt. IgM-Antikörper werden zeitlich vor den IgG-Antikörper gebildet. Kann man aufgrund des Wissens um die Existenz der natürlichen Antikörper (Abschnitt 1.2.3) davon ausgehen, dass alle IgG-Antikörper einen IgM-Vorläufer haben? Falls dem so wäre, könnte das den Unterschied der Balkenhöhen in den Abbildungen 6.5(b) und 6.6(b) erklären. In letzterer ist der rote Balken (früh nach Infektion) für die IgG-Antikörper-Signale am höchsten, während die Signale der IgM-Antikörper einen fortgeschrittenen Verlauf zeigen: Der Effekt der Verschiebung ist bereits beim

blauen Balken (Gruppe »spät«) sichtbar. Der kleine Unterschied in Abbildung 6.5 zwischen dem schwarzen (gesund oder 0 Tage nach Infektion) und dem roten Balken (früh nach Infektion) würde demnach bedeuten, dass im IgM-Repertoire zu einem frühen Zeitpunkt nach Infektion die Spitze der Antwort bereits wieder abgeklungen ist.

Die Tatsache, dass Peptide sowohl in IgM- wie auch in IgG-Detektionen vorkommen (Tabelle 6.2) ruft den Verdacht einer Kreuzreaktivität der sekundären Antikörper hervor. Diesen Verdacht können wir mit den vorliegenden Messungen leider nicht entkräften.

### 8.7.3 IgM- und IgG-Antikörperbindungsmessungen im Vergleich

Wir haben in Kapitel 5 ausschliesslich IgM-Ak-Reaktivitäten gemessen; einerseits aufgrund der Einschränkung auf eine Klasse, die durch die Zellulosemembran-Technologie gegeben war, andererseits, weil uns die IgM-Ak-Repertoires interessanter schienen. Die Resultate der IgM- und IgG-Ak-Reaktivitäten in Kapitel 6 geben uns nachträglich recht (unabhängig von eventuellen Kreuzreaktivitäten): Die Tatsache, dass die SI der IgG-Antikörper-Bindungen, die auch in den meisten käuflichen, diagnostischen Antikörper-tests untersucht werden, in der Lage sind, Unterschiede zwischen den analysierten Gruppen zu finden, ist weniger erstaunlich als die Gegebenheit, dass die IgM-Antikörper eine erhöhte Unterscheidungsfähigkeit aufweisen (in Kapitel 6). Das IgM-Antikörperrepertoire, als Teil des natürlichen Antikörper-repertoires, wurde in den letzten Jahren wenig beachtet. Das Interesse an den natürlichen Antikörpern kann aber bis weit ins letzte Jahrhundert zurückverfolgt werden [33], [11]. Die höchsten diagnostischen Zuverlässigkeiten (DZ) in Tabelle 6.1 stammen von den IgM-Ak-Bindungs-Differenzen zwischen den gesunden und den mit *H.-polygyrus*-infizierten Mäusen früh und spät nach Infektion. Auch der Krankheitsverlauf in Abbildung 6.10(a) ist sehr klar im IgM-Repertoire sichtbar; die Trennung zwischen gesund und infiziert wird im IgM-Repertoire sogar deutlicher als im IgG-Repertoire (Abbildung 6.10(b)). Das Ergebnis der Differenzen im IgM-Repertoire aufgrund der *H.-polygyrus*-Infektion ist eine interessante Entdeckung. Es ist bekannt, dass eine *H.-polygyrus*-Infektion bei BALB/c-Mäusen eine T-Helferzellen-Antwort auslöst, was die Immunantwort in Richtung humoraler Reaktion treibt. Diese humorale Reaktion ist gekennzeichnet durch IL-4-Produktion [46] und einer Steuerung der Immunantwort Richtung IgG1-, IgE- und IgA-Antworten [8], [7] mit vor allem hohen IgG1-Antikörper-Titer [4], [8]. Die

Änderung des IgM-Antikörper-Titers aufgrund von *H.-polygyrus*-Infektionen ist jedoch bisher nicht publiziert und wird in unseren Arbeiten das erste Mal dokumentiert. Der Nachweis, dass *H.-polygyrus*-Infektionen das IgM Repertoire ändern, wirft auch weitere Fragen bezüglich der Effektor-Funktionen verschiedener Antikörperklassen und -unterklassen auf.

## 8.8 Shape Space und Modellierung

Die Reduktion des Problems der Klassifizierung auf wenige Merkmale entspricht dem Auffinden von Lösungen in wenig Dimensionen. Die Idee der Dimensionsreduktion wiederum passt zur Theorie des Shape-Space-Modells von Perelson u. Oster [54]. Gemäss ihren theoretischen Überlegungen sind alle Antikörper-Antigen-Interaktionen in fünf- bis zehndimensionalen »Räumen« repräsentierbar. Zusammenfassend basiert das klassische Shape-Space-Modell auf Punkten im Raum, deren Abstände die Affinitäten zwischen Antikörpern und Antigenen repräsentieren (euklidischer Shape Space). Eine Beschreibung findet sich in Abschnitt 1.8.1. Die Grössenordnung der Dimensionen hängt nicht nur von den chemischen Eigenschaften und der Diversität der Moleküle ab, sondern auch von der Definition des »Raumes«. In der Publikation von Smith u. a. [70] werden Shape-Space-Parameter aus immunologischen Daten abgeleitet. Die Autoren schätzen die Zahl der benötigten Dimensionen zur Beschreibung der Kreuzreaktivität von experimentell gewonnenen Daten auf 20 bis 25 für den Hamming-Shape-Space und auf 5 bis 8 Dimensionen für den Euklidischen-Shape-Space. Im Hamming-Shape-Space sind Antikörper und Antigene als Sequenzen (oder Strings) von Symbolen (oder Buchstaben und Zahlen) dargestellt. Aufgrund der ursprünglichen Definition in der Codierungstheorie entspricht die Hammingdistanz zwischen zwei Codewörtern  $a$  und  $b$  der Anzahl der Positionen in den zwei Sequenzen gleicher Länge, die sich unterscheiden:

$$d(a, b) = |\{i | a_i \neq b_i\}| \quad (8.1)$$

In biologischer Umgebung wird diese Definition, wie z. B. in der Modellbeschreibung von Farmer u. a. [16] abgewandelt in die Anzahl Positionen von zwei Strings, die nicht komplementär zueinander sind. Im einfachsten Fall werden dafür Antikörper und Antigen als binäre Strings beschrieben, die nur aus zwei unterschiedlichen Aminosäuren bestehen und jeweils nur ein Epitop enthalten. Die Hammingdistanz zwischen zwei Strings ( $d(a, b)$ ) entspricht der Anzahl Positionen, die man bei String  $a$  ändern muss, um ihn in den komplementären String  $b$  zu überführen. Die Grösse der möglichen Hammingdistanz hängt von der Zahl der möglichen Buchstaben des String-Alphabets und von



der Länge  $B$  der Strings ab. Es liegt nahe, die Behauptung aufzustellen, dass jegliche Komplexität des String-Alphabets, die das Binäre übersteigt, durch eine Verlängerung des Vektors repräsentiert werden kann.

Zusammenfassend können wir festhalten, dass, wie bereits in Abschnitt 1.8.1 erwähnt, die Literatur [6], [39] von mindestens fünf Dimensionen ausgeht, die ein optimaler Shape Space aufweisen muss. Die Resultate dieser Arbeit und das hohe Differenzierungsvermögen mit den wenigen Peptiden korrelieren gut mit den wenigen Dimensionen des Shape Space. Alle untersuchten Gruppen, die aufgrund ihrer SI-Unterschiede signifikant klassifizierbar sind, lassen sich mittels fünf oder weniger Peptiden unterscheiden. Der Grenzwert der Nachweisbarkeit der Antikörper-Antigen-Wechselwirkung wird durch die Sensitivität der Technologie vorgegeben. Die Anzahl Antigene, die einen einzelnen Antikörper erkennen, und umgekehrt, hat daher keinen erheblichen Einfluss auf die Modellvorstellung des Shape Space, ist auf theoretischer Ebene aber gleichwohl ein wichtiger Parameter.

Ideale experimentelle Messungen zur Modellierung und Abschätzung eines Shape Space wären z. B. Messungen von Bindungen verschiedenster monoklonaler Antikörper gegenüber grossen »idealen« Probenbibliotheken. Die Kenntnis der Verwandtschaft der Antikörper untereinander, z. B. anhand der Ähnlichkeit ihrer CDR-Regionen (siehe Abschnitt 1.2.2), könnte den Zusammenhang zwischen Struktur und Affinität erleichtern, und damit die potenzielle Vorhersage verbessern. Die detaillierte Analyse des Bindungsverhaltens einzelner Antikörper oder gar komplexer Antikörpermischungen, wie sie in Seren vorhanden sind, ist bisher nicht bewältigt. Das experimentelle Unvermögen liegt darin begründet, dass zu einer detaillierten Charakterisierung der Antikörpermischungen die Affinitätsmessungen zu möglichst vielen und diversen *einzelnen* Bindungspartnern (mindestens hunderte, im Idealfall tausende, um ein möglichst repräsentatives Spektrum des möglichen Bindungspartner-Raumes abzudecken) nötig wären, was selbst mit Hilfe von Hochdurchsatz-Messmethoden aufwendig bleibt. Die Verfügbarkeit oder Herstellung möglichst gut charakterisierter Antikörper ist ein unter idealen Umständen lösbares, aber ebenfalls aufwendiges Problem. Um von aufwendigen Bindungsmessung von vielen, einzelnen monoklonalen Antikörpern wegzukommen, können Seren als Mischungen von Antikörpern betrachtet werden. Diese Arbeit beantwortet die Frage, ob das unterschiedliche funktionelle Bindungsverhalten der Antikörper von Serengruppen ebenfalls in fünf Dimensionen dargestellt werden kann, zumindest für die betrachteten Fälle. Allerdings sind bei der Analyse von Antikörpermischungen Eigenschaften wie Diversität und Spezifität von einzelnen Antikörpern sowie deren Mengenanteil und ihr Beitrag

zum Bindungsmuster des Gesamtgemisches keine direkt extrahierbaren Parameter. Derartige Eigenschaften und ihre Kreuzvergleiche wären aber von ausschlaggebendem Interesse, um die Resultate der Peptidarray-Messung zu verstehen und einen Shape Space konstruieren zu können.

Wegen des Mangels an experimentellen Daten der Bindungsmessungen vieler monoklonaler Antikörper gegenüber identischer Proben verfolgen wir die weiteren Shape-Space-Überlegungen vorerst theoretisch anhand von Vektoren und Bitstrings. Anhand einfacher Modelle könnte ein Shape Space konstruiert werden, mit dem die Eigenschaften der differenziellen Antigenerkennung verschiedener Antikörper und ihrer Mutationen abgebildet werden könnten.

## 8.9 Mathematische Modellierung der Antikörpererkennung

Können Modellierungen und Computersimulationen bei der Untersuchung des Bindungsverhaltens von Serumantikörpern an Peptidarrays helfen? Und wenn ja, wie können die Faktoren des Bindungsverhaltens, das von Diversität, Spezifität und Konzentration der einzelnen Klone abhängt, simuliert werden? Ein gutes Computermodell könnte experimentelle Beobachtung einbinden, um statistische Eigenschaften zu inferieren. Im Idealfall würde aus der Simulation eine Methode zur modellbasierten Datenanalyse resultieren. Modellbasierte Datenanalyse würde bedeuten, dass experimentell gewonnene Daten (wie z. B. Peptidarray-Bindungs-Ergebnisse) anhand eines mathematischen Modells interpretiert werden könnten, um zugrunde liegende Phänomene zu beschreiben. Die Grundzüge eines solchen Modells werden im folgenden Abschnitt besprochen und mögliche Schlussfolgerungen diskutiert.

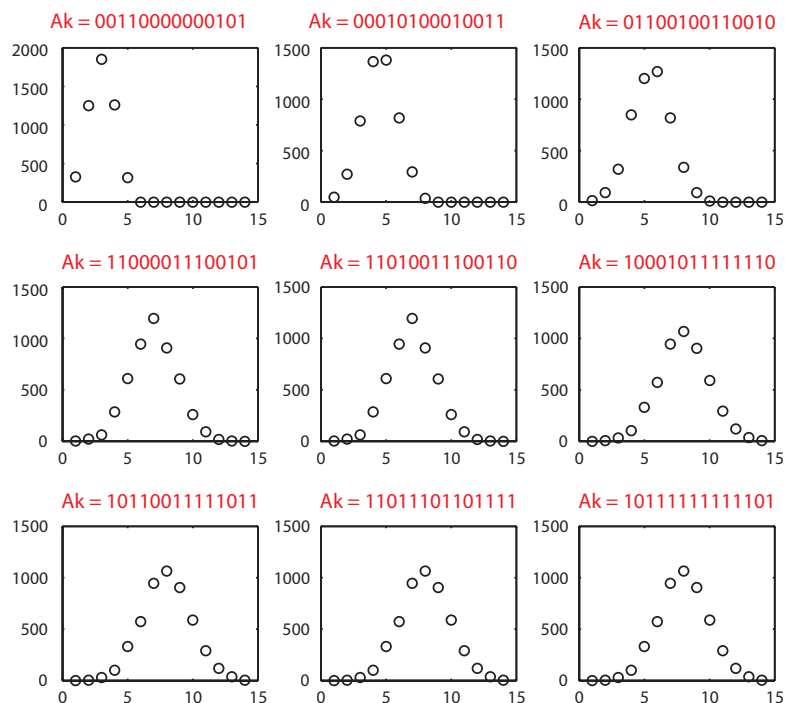
### 8.9.1 Das Bitstring-Modell von Lancet

Ein Bitstring-Modell dient der Simulation von Antikörper-Antigen-Interaktionen in detaillierter Weise. Das *RAD* (von engl. *receptor affinity distribution*)-Bitstring-Modell von Lancet u. a. [37] und Rosenwald u. a. [65] hat den Vorteil, dass es sowohl einfach, als auch bereits besprochen und etabliert ist. Daher liegt es nahe, das *RAD*-Modell als Grundlage zu nutzen. Das Modell basiert auf Bitstrings der Länge  $B$ , die sowohl Antikörper als auch Antigene darstellen und aus mindestens zwei unterschiedlichen Bits bestehen (z. B. 0 und 1). Aus den Differenzen der Bitstrings wird die Affinität aufgrund des

von Farmer u. a. [16] beschriebenen Modells als Hammingdistanz berechnet.

Ein weiterer Vorteil des Bitstring-Modells ist die Möglichkeit der schrittweisen Erhöhung der Dimensionen, indem zuerst das Bindungsverhalten von einzelnen, monoklonalen, theoretischen Antikörpern mit beliebigen Eigenschaften gegen tausende von unterschiedlichen Antigenen untersucht wird. Mit Mischungen solcher Antikörper könnten dann Serien modelliert werden, wobei deren detaillierte Spezifikationen wie Anteile der einzelnen Klone, Diversität, Repertoiregrösse und Spezifität beliebig variiert und untersucht werden könnten. Abbildung 8.1 zeigt als Beispiel die Verteilung der Affinitäten von neun »Antikörper«-Strings gegenüber der gleichen Probenbibliothek, die aus 5000 verschiedenen, zufällig generierten »Antigen«-Strings besteht. Die durchschnittliche Affinität steigt mit zunehmender Anzahl von 1en (von vier bis zwölf). Die Verteilung der Affinität lässt sich demnach alleine durch die Anzahl der 1en im Antikörper-String steuern. Die Positionen der 1en sind dabei für die Histogramme der Affinitäten nicht von Bedeutung; allein die Anzahl der 1en steuert die Verteilung. Dieser Effekt beruht auf der Grösse und Diversität der Probenbibliothek. Die Spezifität wurde in Abschnitt 1.2.6 gegenüber einzelnen Antigenen definiert. Mit diesem Modell, bzw. durch die Betrachtung der Affinitätsverteilung im Histogramm, gibt es keine Möglichkeit zur Änderung der Spezifität gegenüber der gesamten Probenbibliothek; dazu wäre ein weiterer Buchstaben im Bitstring-Alphabet oder eine Einteilung des Bitstrings in unterschiedliche Abschnitte mit unterschiedlichen Bedeutungen nötig. Ein binäres String-Alphabet ist im vorliegenden Fall nicht ausreichend, im Gegensatz zur vorhergehend Überlegung (Abschnitt 8.8).

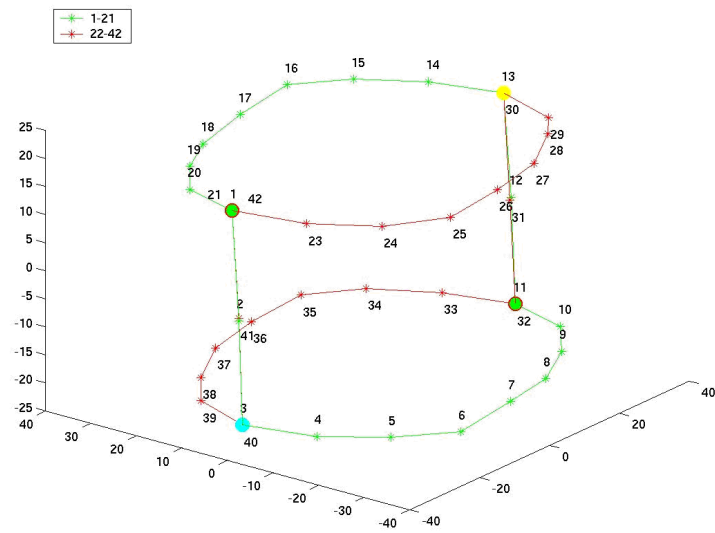
Um vom Bitstring-Modell auf die Dimensionsreduktion zurückzugelangen, kann mittels Berechnung der Hammingdistanzen auch die Bit-für-Bit-Veränderung eines Bitstring-Antikörpers und die Auswirkung dieser Veränderung auf die Affinitätsmessung gegenüber einer Bibliothek von 5000 Peptiden betrachtet werden. Tabelle 8.1 zeigt einen Antikörper, der auf zwei verschiedene Arten Schritt für Schritt von einem Bitstring aus lauter 0en in einen Bitstring aus lauter 1en und wieder zurückmutiert wurde. Identische Antikörper, die dabei entstehen, sind mit der gleichen Farbe gekennzeichnet. Die Affinitätsmatrix dieser Antikörper gegenüber 5000 zufälligen Antigenen wurde berechnet, mittels multidimensionaler Skalierung auf drei Dimensionen reduziert und in Abbildung 8.2 dargestellt. Die Darstellung veranschaulicht, dass die Affinitätsmessungen eines Bitstring-Antikörpers der Länge  $B = 10$ , der Bit für Bit umgewandelt wird, gemessen, wie oben beschrieben, an 5000 zufälligen Antikörpern, in drei Dimensionen dargestellt werden kann. Die Darstellung der Dimensionsreduktion von Affinitätsmessungen funktioniert



**Abbildung 8.1:** Die simulierte Verteilung der Affinitäten von neun verschiedenen Bitstring-Antikörpern mit verschiedener Anzahl von 1en. Der erste Antikörper enthält vier 1en, der letzte zwölf. Die Affinitäten wurden gegenüber 5000 verschiedenen Bitstring-Antigenen (d.h. Peptide) gemessen.

und scheint sinnvoll. Eine Darstellung der Modellvorstellung in einem metrischen Raum scheint im Rahmen dieses sehr einfachen Modells ebenfalls zu funktionieren. Wie sehr diese Modellansätze auf komplexere Gemische und Fragestellungen anwendbar sind, wird sich zeigen. Die Beschreibung des effektiven chemischen Ak-Ag-Raumes wird sich kaum auf zehn binäre Deskriptoren reduzieren lassen.

Mit Modellen dieser Art könnten aber nebst der Frage nach der Dimension des chemischen Antikörper-Antigen-Interaktion-Raumes auch die Frage nach der idealen Bibliothek angegangen werden und Vorschläge für neue Bibliothekdesigns erörtert werden, wie Redestig u. a. [59] detaillierter aufzeigen wird.



**Abbildung 8.2:** Die simulierten Affinitäten einer Schritt-für-Schritt Antikörper-Mutation gemessen an jeweils 5000 Antikörpern im dreidimensionalen Raum. Die Affinitäten der Antikörper aus Tabelle 8.1 gemessen an 5000 Antigenen sind mittels multidimensionaler Skalierung in drei Dimensionen darstellbar.

1:	0 0 0 0 0 0 0 0 0 0	22:	0 0 0 0 0 0 0 0 0 0
2:	1 0 0 0 0 0 0 0 0 0	23:	0 0 1 0 0 0 0 0 0 0
3:	1 1 0 0 0 0 0 0 0 0	24:	0 0 1 1 0 0 0 0 0 0
4:	1 1 1 0 0 0 0 0 0 0	25:	0 0 1 1 1 0 0 0 0 0
5:	1 1 1 1 0 0 0 0 0 0	26:	0 0 1 1 1 1 0 0 0 0
6:	1 1 1 1 1 0 0 0 0 0	27:	0 0 1 1 1 1 1 0 0 0
7:	1 1 1 1 1 1 0 0 0 0	28:	0 0 1 1 1 1 1 1 0 0
8:	1 1 1 1 1 1 1 0 0 0	29:	0 0 1 1 1 1 1 1 1 0
9:	1 1 1 1 1 1 1 1 0 0	30:	0 0 1 1 1 1 1 1 1 1
10:	1 1 1 1 1 1 1 1 1 0	31:	1 0 1 1 1 1 1 1 1 1
11:	1 1 1 1 1 1 1 1 1 1	32:	1 1 1 1 1 1 1 1 1 1
12:	0 1 1 1 1 1 1 1 1 1	33:	1 1 0 1 1 1 1 1 1 1
13:	0 0 1 1 1 1 1 1 1 1	34:	1 1 0 0 1 1 1 1 1 1
14:	0 0 0 1 1 1 1 1 1 1	35:	1 1 0 0 0 1 1 1 1 1
15:	0 0 0 0 1 1 1 1 1 1	36:	1 1 0 0 0 0 1 1 1 1
16:	0 0 0 0 0 1 1 1 1 1	37:	1 1 0 0 0 0 0 1 1 1
17:	0 0 0 0 0 0 1 1 1 1	38:	1 1 0 0 0 0 0 0 1 1
18:	0 0 0 0 0 0 0 1 1 1	39:	1 1 0 0 0 0 0 0 0 1
19:	0 0 0 0 0 0 0 0 1 1	40:	1 1 0 0 0 0 0 0 0 0
20:	0 0 0 0 0 0 0 0 0 1	41:	0 1 0 0 0 0 0 0 0 0
21:	0 0 0 0 0 0 0 0 0 0	42:	0 0 0 0 0 0 0 0 0 0

**Tabelle 8.1:** Die Bitstring-Antikörper, die mittels Bit-für-Bit-Mutationen schrittweise verändert an einer 5000 Antigen-Probenbibliotheken getestet worden sind.

# Literaturverzeichnis

- [1] ABBAS, A. K. ; LICHTMAN, H. L. ; POBER, J. S.: *Cellular and molecular immunology*. 4th. W. B. Saunders company, 2000
- [2] ALT, F. W. ; BLACKWELL, T. K. ; YANCOPOULOS, G. D.: Development of the primary antibody repertoire. In: *Science* 238 (1987), Nr. 4830, S. 1079–87. – 0036-8075 (Print) Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. Review
- [3] ANDRESEN, H. ; GROTZINGER, C. ; ZARSE, K. ; KREUZER, O. J. ; EHRENTREICH-FORSTER, E. ; BIER, F. F.: Functional peptide microarrays for specific and sensitive antibody diagnostics. In: *Proteomics* 6 (2006), Nr. 5, S. 1376–84. – 1615-9853 (Print) Journal Article Research Support, Non-U.S. Gov't
- [4] ANTHONY, R. M. ; RUTITZKY, L. I. ; URBAN, Jr. J. F. J. F. ; STADECKER, M. J. ; GAUSE, W. C.: Protective immune mechanisms in helminth infection. In: *Nat Rev Immunol* 7 (2007), Nr. 12, S. 975–87. – 1474-1741 (Electronic) Journal Article Research Support, N.I.H., Extramural Review
- [5] AVRAMEAS, S.: Natural autoantibodies: from 'horror autotoxicus' to 'gnothi seauton'. In: *Immunol Today* 12 (1991), Nr. 5, S. 154–9. – 0167-5699 (Print) Journal Article Review
- [6] B-RAO, C. ; STEWART, J.: Inverse analysis of empirical matrices of idiotypic network interactions. In: *Bull Math Biol* 58 (1996), Nr. 6, S. 1123–53. – 0092-8240 (Print) Journal Article
- [7] BEHNKE, J. M. ; LOWE, A. ; CLIFFORD, S. ; WAKELIN, D.: Cellular and serological responses in resistant and susceptible mice exposed to repeated infection with *Heligmosomoides polygyrus bakeri*. In: *Parasite Immunol* 25 (2003), Nr. 6, S. 333–40. – 0141-9838 (Print) Journal Article Research Support, Non-U.S. Gov't

- [8] BEN-SMITH, A. ; WAHID, F. N. ; LAMMAS, D. A. ; BEHNKE, J. M.: The relationship between circulating and intestinal Heligmosomoides polygyrus-specific IgG1 and IgA and resistance to primary infection. In: *Parasite Immunol* 21 (1999), Nr. 8, S. 383–95. – 0141-9838 (Print) Journal Article
- [9] BEREK, C. ; MILSTEIN, C.: The dynamic nature of the antibody repertoire. In: *Immunol Rev* 105 (1988), S. 5–26. – 0105-2896 (Print) Journal Article Review
- [10] BONGARTZ, J. ; BRUNI, N. ; OR-GUIL, M.: Epitope Mapping Using Randomly Generated Peptide Libraries. In: REINEKE, U. (Hrsg.): *Epitope Mapping Protocols: Second Edition (Methods in Molecular Biology)*. Totowa, NJ : Humana Press; 2 edition (June 1, 2008), 2008
- [11] BOYDEN, S. V.: Natural antibodies and the immune response. In: *Adv Immunol* 5 (1966), S. 1–28. – 0065-2776 (Print) Journal Article Review
- [12] BRISSAC, C. ; NOBREGA, A. ; CARNEIRO, J. ; STEWART, J.: Functional diversity of natural IgM. In: *Int Immunol* 11 (1999), Nr. 9, S. 1501–7. – 0953-8178 Journal Article
- [13] CHATTERJEE, M. ; MOHAPATRA, S. ; IONAN, A. ; BAWA, G. ; ALI-FEHMI, R. ; WANG, X. ; NOWAK, J. ; YE, B. ; NAHHAS, F. A. ; LU, K. ; WITKIN, S. S. ; FISHMAN, D. ; MUNKARAH, A. ; MORRIS, R. ; LEVIN, N. K. ; SHIRLEY, N. N. ; TROMP, G. ; ABRAMS, J. ; DRAGHICI, S. ; TAINSKY, M. A.: Diagnostic markers of ovarian cancer by high-throughput antigen cloning and detection on arrays. In: *Cancer Res* 66 (2006), Nr. 2, S. 1181–90. – 0008-5472 (Print) Evaluation Studies Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.
- [14] COHN, M. ; LANGMAN, R. E.: The protecton: the unit of humoral immunity selected by evolution. In: *Immunol Rev* 115 (1990), S. 11–147. – 0105-2896 (Print) Journal Article Research Support, U.S. Gov't, P.H.S. Review
- [15] COLECLOUGH, C.: A critique of the Cohn-Langman protection theory. In: *Immunol Rev* 115 (1990), S. 173–81; discussion 182–9. – 0105-2896 (Print) Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. Review



- [16] FARMER, J D. ; PACKARD, N H. ; PERELSON, A S.: The immune system, adaptation, and machine learning. In: *Phys. D* 2 (1986), Nr. 1-3, S. 187–204. – ISSN 0167–2789
- [17] FRANK, R.: The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports—principles and applications. In: *J Immunol Methods* 267 (2002), Nr. 1, S. 13–26. – 0022-1759 Journal Article Review Review, Tutorial
- [18] FRANK, R. ; DORING, R.: Simultaneous Multiple Peptide-Synthesis under Continuous-Flow Conditions on Cellulose Paper Disks as Segmental Solid Supports. In: *Tetrahedron* 44 (1988), Nr. 19, S. 6031–6040. – Article Q6007 English Times Cited:156 Cited References Count:19
- [19] FRANK, Steven A.: *Immunology and evolution of infectious disease*. Princeton, N.J. : Princeton University Press, 2002. – Steven A. Frank
- [20] GEYSEN, H. M. ; MELOEN, R. H. ; BARTELING, S. J.: Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. In: *Proc Natl Acad Sci U S A* 81 (1984), Nr. 13, S. 3998–4002. – 0027-8424 (Print) Journal Article
- [21] GIBBONS, Jean D.: *Nonparametric statistical inference*. 2nd. New York : M. Dekker, 1985. – Jean Dickinson Gibbons. ill. ; 24 cm. Includes index.
- [22] HAURY, M. ; GRANDIEN, A. ; SUNDBLAD, A. ; COUTINHO, A. ; NOBREGA, A.: Global analysis of antibody repertoires. 1. An immunoblot method for the quantitative screening of a large number of reactivities. In: *Scand J Immunol* 39 (1994), Nr. 1, S. 79–87. – 0300-9475 Journal Article
- [23] HAYAKAWA, K. ; HARDY, R. R. ; HONDA, M. ; HERZENBERG, L. A. ; STEINBERG, A. D. ; HERZENBERG, L. A.: Ly-1 B cells: functionally distinct lymphocytes that secrete IgM autoantibodies. In: *Proc Natl Acad Sci U S A* 81 (1984), Nr. 8, S. 2494–8. – 0027-8424 (Print) Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.
- [24] HILLER, R. ; LAFFER, S. ; HARWANEGG, C. ; HUBER, M. ; SCHMIDT, W. M. ; TWARDOSZ, A. ; BARLETTA, B. ; BECKER, W. M. ; BLASER, K. ; BREITENEDER, H. ; CHAPMAN, M. ; CRAMERI, R. ; DUCHENE, M. ; FERREIRA, F. ; FIEBIG, H. ; HOFFMANN-SOMMERGRUBER, K. ; KING,

- T. P. ; KLEBER-JANKE, T. ; KURUP, V. P. ; LEHRER, S. B. ; LIDHOLM, J. ; MULLER, U. ; PINI, C. ; REESE, G. ; SCHEINER, O. ; SCHEYNIUS, A. ; SHEN, H. D. ; SPITZAUER, S. ; SUCK, R. ; SWOBODA, I. ; THOMAS, W. ; TINGHINO, R. ; VAN HAGE-HAMSTEN, M. ; VIRTANEN, T. ; KRAFT, D. ; MULLER, M. W. ; VALENTA, R.: Microarrayed allergen molecules: diagnostic gatekeepers for allergy treatment. In: *Faseb J* 16 (2002), Nr. 3, S. 414–6. – 1530-6860 (Electronic) Journal Article
- [25] HOCHREITER, S. ; OBERMAYER, K.: Nonlinear Feature Selection with the Potential Support Vector Machine. In: GUYON, I. (Hrsg.) ; GUNN, S. (Hrsg.) ; NIKRAVESH, M. (Hrsg.) ; ZADEH, L. (Hrsg.): *Feature Extraction: Foundations and Applications*. Berlin; New York : Springer Verlag, 2006, S. 419–438
- [26] HOCHREITER, S. ; OBERMAYER, K.: Support vector machines for dyadic data. In: *Neural Comput* 18 (2006), Nr. 6, S. 1472–510. – 0899-7667 (Print) Journal Article
- [27] HOLLANDER, Myles ; WOLFE, Douglas A.: *Nonparametric statistical methods*. New York, : Wiley, 1973 (Wiley series in probability and mathematical statistics). – [by] Myles Hollander [and] Douglas A. Wolfe. illus. 24 cm.
- [28] HORN, S. ; LUEKING, A. ; MURPHY, D. ; STAUDT, A. ; GUTJAHR, C. ; SCHULTE, K. ; KONIG, A. ; LANDSBERGER, M. ; LEHRACH, H. ; FELIX, S. B. ; CAHILL, D. J.: Profiling humoral autoimmune repertoire of dilated cardiomyopathy (DCM) patients and development of a disease-associated protein chip. In: *Proteomics* 6 (2006), Nr. 2, S. 605–13. – 1615-9853 (Print) Journal Article Research Support, Non-U.S. Gov't
- [29] HUEBER, W. ; KIDD, B. A. ; TOMOOKA, B. H. ; LEE, B. J. ; BRUCE, B. ; FRIES, J. F. ; SONDERSTRUP, G. ; MONACH, P. ; DRIJFHOUT, J. W. ; VENROOIJ, W. J. ; UTZ, P. J. ; GENOVESE, M. C. ; ROBINSON, W. H.: Antigen microarray profiling of autoantibodies in rheumatoid arthritis. In: *Arthritis Rheum* 52 (2005), Nr. 9, S. 2645–55. – 0004-3591 (Print) Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.
- [30] IGNATOVICH, O. ; TOMLINSON, I. M. ; POPOV, A. V. ; BRUGGEMANN, M. ; WINTER, G.: Dominance of intrinsic genetic factors in shaping the human immunoglobulin V $\lambda$  repertoire. In: *J Mol Biol* 294 (1999), Nr. 2, S. 457–65. – 0022-2836 (Print) Journal Article Research Support, Non-U.S. Gov't

- [31] JACKSON, J. E.: *A user's guide to principal components*. Hoboken, N.J. : Wiley, 1991 (Wiley series in probability and statistics). – J. Edward Jackson. ill. ; 24 cm. "Wiley-Interscience paperback series-Cover P.[4]. Errata inserted. 1. Getting Started – 2. PCA With More Than Two Variables – 3. Scaling of Data – 4. Inferential Procedures – 5. Putting It All Together - Hearing Loss I – 6. Operations with Group Data – 7. Vector Interpretation I: Simplifications and Inferential Techniques – 8. Vector Interpretation II: Rotation – 9. A Case History - Hearing Loss II – 10. Singular Value Decomposition: Multidimensional Scaling I – 11. Distance Models: Multidimensional Scaling II – 12. Linear Models I: Regression; PCA of Predictor Variables – 13. Linear Models II: Analysis of Variance; PCA of Response Variables – 14. Other Applications of PCA – 15. Flatland: Special Procedures for Two Dimensions – 16. Odds and Ends – 17. What is Factor Analysis Anyhow? – 18. Other Competitors – App. A. Matrix Properties – App. B. Matrix Algebra Associated with Principal Component Analysis – App. C. Computational Methods – App. D. A Directory of Symbols and Definitions for PCA – App. F. Data Sets Used in This Book.
  
- [32] JANEWAY, Charles A.: *Immunologie*. 5. Aufl. Heidelberg : Spektrum Akademischer Verlag, 2002. – Charles A. Janeway jr. ... [ et al.] ; aus dem Engl. Übers. von Kurt Beigeborn ... [et al.] ; mit einem Vorw. zur dt. Ausg. von Klaus Rajewsky Ill., graph. Darst.
  
- [33] JERNE, N. K.: The Natural-Selection Theory of Antibody Formation. In: *Proc Natl Acad Sci U S A* 41 (1955), Nr. 11, S. 849–57. – 0027-8424 (Print) Journal Article
  
- [34] KNEBEL, T. ; HOCHREITER, S. ; OBERMAYER, K.: An SMO algorithm for the potential support vector machine. In: *Neural Computation* 20 (2008), Nr. 1, S. 271–287. – Article 240UK English Times Cited:0 Cited References Count:13
  
- [35] KOHLER, G. ; MILSTEIN, C.: Continuous cultures of fused cells secreting antibody of predefined specificity. In: *Nature* 256 (1975), Nr. 5517, S. 495–7. – 0028-0836 (Print) Journal Article
  
- [36] KRAMER, A. ; REINEKE, U. ; DONG, L. ; HOFFMANN, B. ; HOFFMULLER, U. ; WINKLER, D. ; VOLKMER-ENGERT, R. ; SCHNEIDER-MERGENER, J.: Spot synthesis: observations and optimizations. In: *J Pept Res* 54 (1999), Nr. 4, S. 319–27. – 1397-002x Journal Article

- [37] LANCET, D. ; KEDEM, O. ; PILPEL, Y.: Emergence of order in small autocatalytic sets maintained far from equilibrium: application of a probabilistic receptor affinity distribution (RDA) model. In: *BER BUN-SENSES PHYS CHEM* Bd. 98. Weinheim : Weinheim VCH, 1994, S. 1166–1169
- [38] LANGMAN, R. E. ; COHN, M.: The E-T (elephant-tadpole) paradox necessitates the concept of a unit of B-cell function: the protection. In: *Mol Immunol* 24 (1987), Nr. 7, S. 675–97. – 0161-5890 (Print) Journal Article Research Support, U.S. Gov't, P.H.S.
- [39] LAPEDES, A. ; FARBER, R.: The geometry of shape space: application to influenza. In: *J Theor Biol* 212 (2001), Nr. 1, S. 57–69. – 0022-5193 (Print) Journal Article
- [40] LI, Q. Z. ; XIE, C. ; WU, T. ; MACKAY, M. ; ARANOW, C. ; PUTTERMAN, C. ; MOHAN, C.: Identification of autoantibody clusters that best predict lupus disease activity using glomerular proteome arrays. In: *J Clin Invest* 115 (2005), Nr. 12, S. 3428–39. – 0021-9738 (Print) Journal Article Research Support, N.I.H., Extramural
- [41] LIU, Y. J.: Sites of B lymphocyte selection, activation, and tolerance in spleen. In: *J Exp Med* 186 (1997), Nr. 5, S. 625–9. – 0022-1007 (Print) Journal Article Review
- [42] MAIZELS, R. M. ; YAZDANBAKHS, M.: Immune regulation by helminth parasites: cellular and molecular mechanisms. In: *Nat Rev Immunol* 3 (2003), Nr. 9, S. 733–44. – 1474-1733 (Print) Journal Article Research Support, Non-U.S. Gov't Review
- [43] MATTHEAKIS, L. C. ; BHATT, R. R. ; DOWER, W. J.: An in vitro polysome display system for identifying ligands from very large peptide libraries. In: *Proc Natl Acad Sci U S A* 91 (1994), Nr. 19, S. 9022–6. – 0027-8424 (Print) In Vitro Journal Article
- [44] MERBL, Y. ; ZUCKER-TOLEDANO, M. ; QUINTANA, F. J. ; COHEN, I. R.: Newborn humans manifest autoantibodies to defined self molecules detected by antigen microarray informatics. In: *J Clin Invest* 117 (2007), Nr. 3, S. 712–8. – 0021-9738 (Print) Journal Article Research Support, Non-U.S. Gov't
- [45] MICHAUD, G. A. ; SALCIUS, M. ; ZHOU, F. ; BANGHAM, R. ; BONIN, J. ; GUO, H. ; SNYDER, M. ; PREDKI, P. F. ; SCHWEITZER, B. I.:

- Analyzing antibody specificity with whole proteome microarrays. In: *Nat Biotechnol* 21 (2003), Nr. 12, S. 1509–12. – 1087-0156 (Print) Evaluation Studies Journal Article
- [46] MOHRS, K. ; HARRIS, D. P. ; LUND, F. E. ; MOHRS, M.: Systemic dissemination and persistence of Th2 and type 2 cells in response to infection with a strictly enteric nematode parasite. In: *J Immunol* 175 (2005), Nr. 8, S. 5306–13. – 0022-1767 (Print) Journal Article
  - [47] NIMMERJAHN, F. ; RAVETCH, J. V.: Fcγ receptors: old friends and new family members. In: *Immunity* 24 (2006), Nr. 1, S. 19–28. – 1074-7613 (Print) Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Review
  - [48] NOBREGA, A. ; GRANDIEN, A. ; HAURY, M. ; HECKER, L. ; MALANCHERE, E. ; COUTINHO, A.: Functional diversity and clonal frequencies of reactivity in the available antibody repertoire. In: *Eur J Immunol* 28 (1998), Nr. 4, S. 1204–15. – 0014-2980 Journal Article
  - [49] NOSSAL, G. J.: Antibody production by single cells. In: *Br J Exp Pathol* 39 (1958), Nr. 5, S. 544–51. – 0007-1021 (Print) Journal Article
  - [50] NOSSAL, G. J.: The protection: a view from the bunker. In: *Immunol Rev* 115 (1990), S. 149–66; discussion 167–72. – 0105-2896 (Print) Journal Article Review
  - [51] OPSTELTEN, D. ; OSMOND, D. G.: Pre-B cells in mouse bone marrow: immunofluorescence stathmokinetic studies of the proliferation of cytoplasmic mu-chain-bearing cells in normal mice. In: *J Immunol* 131 (1983), Nr. 6, S. 2635–40. – 0022-1767 (Print) Journal Article Research Support, Non-U.S. Gov't
  - [52] PARMLEY, S. F. ; SMITH, G. P.: Antibody-selectable filamentous fd phage vectors: affinity purification of target genes. In: *Gene* 73 (1988), Nr. 2, S. 305–18. – 0378-1119 (Print) Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.
  - [53] PEREIRA, P. ; FORNI, L. ; LARSSON, E. L. ; COOPER, M. ; HEUSSER, C. ; COUTINHO, A.: Autonomous activation of B and T cells in antigen-free mice. In: *Eur J Immunol* 16 (1986), Nr. 6, S. 685–8. – 0014-2980 (Print) Journal Article Research Support, Non-U.S. Gov't

- [54] PERELSON, A. S. ; OSTER, G. F.: Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-non-self discrimination. In: *J Theor Biol* 81 (1979), Nr. 4, S. 645–70. – 0022-5193 (Print) Journal Article
- [55] PREDKI, P. F. ; MATTOON, D. ; BANGHAM, R. ; SCHWEITZER, B. ; MICHAUD, G.: Protein microarrays: a new tool for profiling antibody cross-reactivity. In: *Hum Antibodies* 14 (2005), Nr. 1-2, S. 7–15. – 1093-2607 (Print) Journal Article Review
- [56] QUINTANA, F. J. ; COHEN, I. R.: The natural autoantibody repertoire and autoimmune disease. In: *Biomed Pharmacother* 58 (2004), Nr. 5, S. 276–81. – 0753-3322 Journal Article Review Review, Tutorial
- [57] QUINTANA, F. J. ; GETZ, G. ; HED, G. ; DOMANY, E. ; COHEN, I. R.: Cluster analysis of human autoantibody reactivities in health and in type 1 diabetes mellitus: a bio-informatic approach to immune complexity. In: *J Autoimmun* 21 (2003), Nr. 1, S. 65–75. – 0896-8411 (Print) Journal Article
- [58] RADBRUCH, A. ; MUEHLINGHAUS, G. ; LUGER, E. O. ; INAMINE, A. ; SMITH, K. G. ; DORNER, T. ; HIEPE, F.: Competence and competition: the challenge of becoming a long-lived plasma cell. In: *Nat Rev Immunol* 6 (2006), Nr. 10, S. 741–50. – 1474-1733 (Print) Journal Article Research Support, Non-U.S. Gov't Review
- [59] REDESTIG, H. ; BONGARTZ, J. ; BRUNI, N. ; SCHUCHHARDT, J. ; OR-GUIL, M.: Simple correlations from complex mixtures. In: in preparation (2008)
- [60] REINEKE, U. ; IVASCU, C. ; SCHLIEF, M. ; LANDGRAF, C. ; GERICKE, S. ; ZAHN, G. ; HERZEL, H. ; VOLKMER-ENGERT, R. ; SCHNEIDER-MERGENER, J.: Identification of distinct antibody epitopes and mimotopes from a peptide array of 5520 randomly generated sequences. In: *J Immunol Methods* 267 (2002), Nr. 1, S. 37–51. – 0022-1759 Journal Article
- [61] REINEKE, U. ; KRAMER, A. ; SCHNEIDER-MERGENER, J.: Epitope Mapping with Synthetic Peptides Prepared by SPOT Synthesis. In: KONTERMANN, R. (Hrsg.) ; DÜBEL, S. (Hrsg.): *Antibody Engineering*. Berlin; New York : Springer Lab Manual, 2001. – ISBN 3540413545 9783540413547, S. 433–459

- [62] RIPLEY, Brian D.: *Pattern recognition and neural networks*. Cambridge ; New York : Cambridge University Press, 1996. – B.D. Ripley. ill. ; 26 cm.
- [63] ROBINSON, W. H. ; DiGENNARO, C. ; HUEBER, W. ; HAAB, B. B. ; KAMACHI, M. ; DEAN, E. J. ; FOURNEL, S. ; FONG, D. ; GENOVESE, M. C. ; VEGVAR, H. E. ; SKRINER, K. ; HIRSCHBERG, D. L. ; MORRIS, R. I. ; MULLER, S. ; PRUIJN, G. J. ; VENROOIJ, W. J. ; SMOLEN, J. S. ; BROWN, P. O. ; STEINMAN, L. ; UTZ, P. J.: Autoantigen microarrays for multiplex characterization of autoantibody responses. In: *Nat Med* 8 (2002), Nr. 3, S. 295–301. – 1078-8956 (Print) Comparative Study Evaluation Studies Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.
- [64] ROBINSON, W. H. ; FONTOURA, P. ; LEE, B. J. ; VEGVAR, H. E. ; TOM, J. ; PEDOTTI, R. ; DiGENNARO, C. D. ; MITCHELL, D. J. ; FONG, D. ; HO, P. P. ; RUIZ, P. J. ; MAVERAKIS, E. ; STEVENS, D. B. ; BERNARD, C. C. ; MARTIN, R. ; KUCHROO, V. K. ; NOORT, J. M. ; GENAIN, C. P. ; AMOR, S. ; OLSSON, T. ; UTZ, P. J. ; GARREN, H. ; STEINMAN, L.: Protein microarrays guide tolerizing DNA vaccine treatment of autoimmune encephalomyelitis. In: *Nat Biotechnol* 21 (2003), Nr. 9, S. 1033–9. – 1087-0156 (Print) Evaluation Studies Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.
- [65] ROSENWALD, S. ; KAFRI, R. ; LANCET, D.: Test of a statistical model for molecular recognition in biological repertoires. In: *J Theor Biol* 216 (2002), Nr. 3, S. 327–36. – 0022-5193 (Print) Journal Article Research Support, Non-U.S. Gov't
- [66] SAEYS, Y. ; INZA, I. ; LARRANAGA, P.: A review of feature selection techniques in bioinformatics. In: *Bioinformatics* 23 (2007), Nr. 19, S. 2507–17. – 1460-2059 (Electronic) Journal Article Research Support, Non-U.S. Gov't Review
- [67] SANDBERG, M. ; ERIKSSON, L. ; JONSSON, J. ; SJOSTROM, M. ; WOLD, S.: New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. In: *J Med Chem* 41 (1998), Nr. 14, S. 2481–91. – 0022-2623 (Print) Journal Article Research Support, Non-U.S. Gov't
- [68] SHREFFLER, W. G. ; BEYER, K. ; CHU, T. H. ; BURKS, A. W. ; SAMPSON, H. A.: Microarray immunoassay: association of clinical history, in vitro IgE function, and heterogeneity of allergenic peanut epitopes. In: *J*

- Allergy Clin Immunol* 113 (2004), Nr. 4, S. 776–82. – 0091-6749 (Print)  
In Vitro Journal Article Research Support, U.S. Gov't, P.H.S.
- [69] SINGH, M. ; LINGAPPAN, K.: Whooping cough: the current scene. In: *Chest* 130 (2006), Nr. 5, S. 1547–53. – 0012-3692 (Print) Journal Article Review
- [70] SMITH, D. J. ; FORREST, S. ; HIGHTOWER, R. R. ; PERELSON, A. S.: Deriving shape space parameters from immunological data. In: *J Theor Biol* 189 (1997), Nr. 2, S. 141–50. – 0022-5193 (Print) Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.
- [71] STAHL, D. ; LACROIX-DESMAZES, S. ; BARREAU, C. ; SIBROWSKI, W. ; KAZATCHKINE, M. D. ; KAVERI, S. V.: Altered antibody repertoires of plasma IgM and IgG toward nonself antigens in patients with warm autoimmune hemolytic anemia. In: *Hum Immunol* 62 (2001), Nr. 4, S. 348–61. – 0198-8859 Journal Article
- [72] STAHL, D. ; LACROIX-DESMAZES, S. ; BARREAU, C. ; SIBROWSKI, W. ; KAZATCHKINE, M. D. ; KAVERI, S. V.: Altered antibody repertoires of plasma IgM and IgG toward nonself antigens in patients with warm autoimmune hemolytic anemia. In: *Hum Immunol* 62 (2001), Nr. 4, S. 348–61. – 0198-8859 Journal Article
- [73] STAHL, D. ; LACROIX-DESMAZES, S. ; HEUDES, D. ; MOUTHON, L. ; KAVERI, S. V. ; KAZATCHKINE, M. D.: Altered control of self-reactive IgG by autologous IgM in patients with warm autoimmune hemolytic anemia. In: *Blood* 95 (2000), Nr. 1, S. 328–35. – 0006-4971 (Print) Journal Article Research Support, Non-U.S. Gov't
- [74] STAHL, D. ; LACROIX-DESMAZES, S. ; MOUTHON, L. ; KAVERI, S. V. ; KAZATCHKINE, M. D.: Analysis of human self-reactive antibody repertoires by quantitative immunoblotting. In: *J Immunol Methods* 240 (2000), Nr. 1-2, S. 1–14. – 0022-1759 Journal Article Review
- [75] STAHL, D. ; LACROIX-DESMAZES, S. ; SIBROWSKI, W. ; KAZATCHKINE, M. D. ; KAVERI, S. V.: Broad alterations of self-reactive antibody-repertoires of plasma IgM and IgG in B-cell chronic lymphocytic leukemia (B-CLL) and B-CLL related target-restricted autoimmunity. In: *Leuk Lymphoma* 42 (2001), Nr. 1-2, S. 163–76. – 1042-8194 Journal Article



- [76] SUNDBLAD, A. ; FERREIRA, C. ; NOBREGA, A. ; HAURY, M. ; FERREIRA, E. ; PADUA, F. ; COUTINHO, A.: Characteristic generated alterations of autoantibody patterns in idiopathic thrombocytopenic purpura. In: *J Autoimmun* 10 (1997), Nr. 2, S. 193–201. – 0896-8411 Journal Article
- [77] TAPIA, V. ; BONGARTZ, J. ; SCHUTKOWSKI, M. ; BRUNI, N. ; WEISER, A. ; AY, B. ; VOLKMER, R. ; OR-GUIL, M.: Affinity profiling using the peptide microarray technology: a case study. In: *Anal Biochem* 363 (2007), Nr. 1, S. 108–18. – 0003-2697 (Print) Journal Article Research Support, Non-U.S. Gov't
- [78] TONG, M. ; JACOBI, C. E. ; RIJKE, F. M. d. ; KUIJPER, S. ; WERKEN, S. van d. ; LOWARY, T. L. ; HOKKE, C. H. ; APPELMELK, B. J. ; NAGELKERKE, N. J. ; TANKE, H. J. ; GIJLSWIJK, R. P. ; VEUSKENS, J. ; KOLK, A. H. ; RAAP, A. K.: A multiplexed and miniaturized serological tuberculosis assay identifies antigens that discriminate maximally between TB and non-TB sera. In: *J Immunol Methods* 301 (2005), Nr. 1-2, S. 154–63. – 0022-1759 (Print) Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.
- [79] TREPEL, F.: Number and distribution of lymphocytes in man. A critical analysis. In: *Klin Wochenschr* 52 (1974), Nr. 11, S. 511–5. – 0023-2173 (Print) Comparative Study Journal Article
- [80] VASCONCELLOS, R. ; NOBREGA, A. ; HAURY, M. ; VIALE, A. C. ; COUTINHO, A.: Genetic control of natural antibody repertoires: I. IgH, MHC and TCR beta loci. In: *Eur J Immunol* 28 (1998), Nr. 3, S. 1104–15. – 0014-2980 Journal Article
- [81] VEGVAR, H. E. d. ; AMARA, R. R. ; STEINMAN, L. ; UTZ, P. J. ; ROBINSON, H. L. ; ROBINSON, W. H.: Microarray profiling of antibody responses against simian-human immunodeficiency virus: postchallenge convergence of reactivities independent of host histocompatibility type and vaccine regimen. In: *J Virol* 77 (2003), Nr. 20, S. 11125–38. – 0022-538X (Print) Journal Article Research Support, U.S. Gov't, P.H.S.
- [82] VENABLES, W. N. ; RIPLEY, Brian D.: *Modern applied statistics with S*. 4th. New York : Springer, 2002 (Statistics and computing). – W.N. Venables, B.D. Ripley. ill. ; 25 cm. Previously published as: *Modern applied statistics with S-PLUS*, c1999. Typographical Conventions – 1. Introduction – 2. Data Manipulation – 3. The S Language – 4. Graphics – 5. Univariate Statistics – 6. Linear Statistical Models – 7. Generalized

Linear Models – 8. Non-Linear and Smooth Regression – 9. Tree-Based Methods – 10. Random and Mixed Effects – 11. Exploratory Multivariate Analysis – 12. Classification – 13. Survival Analysis – 14. Time Series Analysis – 15. Spatial Statistics – 16. Optimization – App. A. Implementation-Specific Details – App. B. The S-PLUS GUI – App. C. Datasets, Software and Libraries.

- [83] WANG, D. ; LIU, S. ; TRUMMER, B. J. ; DENG, C. ; WANG, A.: Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells. In: *Nat Biotechnol* 20 (2002), Nr. 3, S. 275–81. – 1087-0156 (Print) Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.
- [84] WARDEMAN, H. ; YURASOV, S. ; SCHAEFER, A. ; YOUNG, J. W. ; MEFFRE, E. ; NUSSENZWEIG, M. C.: Predominant autoantibody production by early human B cell precursors. In: *Science* 301 (2003), Nr. 5638, S. 1374–7. – 1095-9203 (Electronic) Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.
- [85] WEISER, A. A. ; OR-GUIL, M. ; TAPIA, V. ; LEICHSENRING, A. ; SCHUCHHARDT, J. ; FROMMEL, C. ; VOLKMER-ENGERT, R.: SPOT synthesis: Reliability of array-based measurement of peptide binding affinity. In: *Anal Biochem* 342 (2005), Nr. 2, S. 300–11. – 0003-2697 Journal Article
- [86] WENSCHUH, H. ; GAUSEPOHL, H. ; GERMEROTH, L. ; ULBRICHT, M. ; MATUSCHEWSKI, H. ; KRAMER, A. ; VOLKMER-ENGERT, R. ; HEINE, N. ; AST, T. ; SCHARN, D. ; SCHNEIDER-MERGENER, J.: *Positionally addressable parallel synthesis on continuous membranes; In Combinatorial Chemistry: A Practical Approach*. Bd. Chap 5. Oxford University Press, Oxford, UK, 2000
- [87] ZHU, H. ; BILGIN, M. ; SNYDER, M.: Proteomics. In: *Annu Rev Biochem* 72 (2003), S. 783–812. – 0066-4154 (Print) Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. Review

# Anhang A

ADTDAGTLTAQGSRH	HEVWTPPLDHFPWHY	QGVRHDLWNDAPSNF
AKKASEPRDNFKALS	HLNMPKSENYCRIRV	QIHAKYVAKFINNLV
ANWTCKNQVIWEEMS	HNTTDTSVKGNMAKD	QKPSASNYWKDWIER
AQDFHRASTNREPDV	HSEVSWKENVPDEDD	QLDTFFSKVWYNLYH
ASRAGPFAAPKVPAD	HSKEWVYECEPQRAA	QLLVEITRLKGHKWY
ATPNFIPLPFYFYGR	IDGSSLSQTYNPEYY	QNDQKREKCRGLPLG
AVRKVRTWRCPVVDG	IESTFVYQENDLRLN	QNKAGWVVELAAKVY
AWFNGWSDFGNTKDF	IHLLEGHFVREEQH	QQSFTQKNNAIQFNK
AYLPVFAHYVDSCDA	IIYKLYRSACGALRD	QSAPLVVPPPLNGIL
CGKSSQYQSDIDLQ	IKFIELTQTARKCAT	QVDEEREGPQLERHC
DAINQRENHEKVEHG	IKHTGMAFKPVEYFW	QVEQVVS RKFGFWKF
DDDKLGKGSPIKSF	ILAEVDHYMRGQDPR	RAHPKWGNVLSFFNA
DDLICFSTVVRPFFG	INTMNWHLGCDFPFL	RALFIFAERDPMKGV
DHCPFMYNRCRDADC	INYHMGDVEKPPGRT	RFYHMF IQKQHQLPW
DIMGWRVTRNYDVHH	IPDNMKTRIVAGIAR	RHRELQIFSERTRHQ
DKELDLEHKPRLHIT	IYRRKKRKQKYNQKV	RMNDPSQVSSFLKKT
DNVQGDFWLNVSYSIS	KFKNHIAGAPYEFYN	RRLSATENELRQGRP
DQMPGLMYHTMAHNF	KFVPWDTKRIFTTVD	RSVWRSFGEHAAWHE
DRHFFQVWNIRQCEN	KGFRKDHSPPFNRND	RTDVLPNYIKAMNGH
DRSRRKAEFMHAFTQ	KKSTHHESWLKIVDY	RTQELYTDPYERITH
DSHWNKRS LGAQPY	KQEYLGLNIAITFQH	RYQQQRTS ISSKWAD
DVPFEFEIDFAQAQN	KQHPEGDQNRSTRGP	RYWVWKILIYYAPHL
DYNCKRMYEVAVCNR	KSAAQSKSKRGGSET	SDQERCEQPDRRIKD
EAFYLRTAHKPVVHR	KVIYDMEYSECYSVG	SDVCSWTSSDMTNMG
EASWDCGDDMDSYWS	KVLRHEHSYTESSRY	SKCQVPTVWIAAFHQ
EDAAHKHHRDLGGQH	KVTPRSNTPLRVQLK	SNVFNSYGPLTTFSH
EEFFLMFEDWKWENL	LGETRNPMMWLVGAG	SPNRLQDGKQWNTSN
EEYRKLT VQWYQPEE	LKAHLSLNLRFNCT	SVDFVAEPVKLPENQ
EFCTVLKMRFPQQWR	LLLMKYAFDDRQAEW	TDCQLWYLQQKIDKF
EFIREINGRHWKTLR	LRFRGTCPMWKCHFL	TEENPTRERPKERG
ENFSDAVYIVRIIIS	LSRRPMHHPDDPHT	TEIEAFLNPRMGQPP

ERSSYHGKYGQSPMS	LVSLASSSARGDTYF	TIDRKWPKTEHTRPD
ESFSWYRDFSRPNIN	MEAGMQDDQFQKHYP	TIPQWWQQDHETAPF
ETQSHFHKGDKVGLK	MPERNSTLPPNYLIS	TQHDNTYRHLAWDYF
EVFDLQTNQDAHQLH	NAPYIIKMAPMVLWA	TTGGPWLAGREWAPG
EYEKNGTRKRPRGAV	NATYDRNGMLWAWTR	VKNRADQTTSPHFWA
FKIIQAKFHPNWDNH	NEPLESAVDVYTKCG	VLTKFENNETDSSNA
FPVHYDFTFELILRM	NFFDEIIHSCRSQNG	VMYRQKQVHTDLFPA
FSGFPDKIEFPTQDC	NLFKTDRGVAHRTMD	VSDPWTWGHHSPLDL
FWKRHYWQTTMPFMT	NPPNEWMDDWQPGEN	VYPKNCIVLLFEWQD
FYFMVIYFALISAYD	NRQLGQPPLLSTFPE	WEHSDWVIFWGIHNT
GAELKLNKDVRGYFF	NTIVNHRGKAENADK	WKRMRTINFRSUNAQ
GFQETGRFHTKSVCA	NTQISFVVPYNSPLS	WLPTATGNMQSVPLI
GHDDFQGPWVPNTE	PAKNENTRYFGNYTG	WSITSFQDMRAPTSQ
GHMKHYEERERVVWI	PDSTSGPVKTKAPVQ	YANHYKDGESEKAN
GINLATSDTEDSPGN	PEPMRHSIRGWNQSQ	YIEKQITNYPGSAKH
GLVTDARTKYKEEGV	PGYLVDHAGLIATY	YNTFLHANCWVAIHG
GRRQIKQTLDSDEVY	PKYTLSHSWHKIAMA	YNYLYNEILSMNKFR
GSHHFNWMNLMAVQY	PSPDMKNTNVAYIGS	YPYFNQLQKDRSDLN
GVGVSTGSYDNQTHY	PSTDDYEDAASQGTE	YRYPANDWRTNSEEI
GVIVPRINEHYHKFN	QEMKDAFGHLHGIRC	YTRRYNIMSKIFNTW
GVVTIKTITKKDMVN	QESQIDQGGWLSEHI	YVRFTLMEITIYTVMS

#### Die 156 Peptide der Zellulosemembranen

QQTWSESDTCDDDY	GLDKIKNNVKAHKL	PTQEPRKVVHNSDEM
QQAKIGKETTIELD	HPTKYTLPKYGVSN	SSYQDKTTHKQYPF
DKRTMDGTYDIKYD	VTDTHSGDSSSNLK	FWDGHRHAMKEMCG
QDDTCETWIEADDN	ALITVHVDSLLWVM	KEKKYDNGPTSRL
DVTDTKLVYATYRE	TERIKSELLGQNTK	AKLRANTTWNVPYQ
GDQNQHVDSEIPWY	AIGKWLD CYGKMPH	GTKDEQDISDFRAA
FKTRVQDGSPWKTA	NKMC GIRGPSGCDE	KCINKTWISGWYD
LARDFRDTDVDSRE	LDKRRQHNYLFVAV	GSPCRRVQDNAAKS
FIREINGRHWKTLR	LLMKYAFDDRQAEW	KELDLEHKPRLHIT
TIVNHRGKAENADK	MYRQKQVHTDLFPA	KFIELTQTARKCAT
VSLASSSARGDTYF	EPLESAVDVYTKCG	SITSFQDMRAPTSQ
ASWDCGDDMDSYWS	FKNHIAGAPYEFYN	FFDEIIHSCRSQNG
YNCKRMYEVAVCNR	EAGMQDDQFQKHYP	GKSSQYQSDIDLQD
ATYDRNGMLWAWTR	EIEAFLNPRMGQPP	SAPLVVPPPLNGIL
GFRKDHSPPFNRND	APYIIKMAPMVLWA	SDPWTWGHHSPLDL
HCPFMYNRCRDADC	TQSHFHKGDKVGLK	SVWRSFGEHAAWHE
YQQQRTSISSKWAD	NYHMGDVEKPPGRT	SAAQSKSKRGGSET

VFDLQTNQDAHQLH	AINQRENHEKVEHG	KYTLSHSHWHKIAMA
IEKQITNYPGAKH	QDFHRASTNREPDV	EEENPTRERPKERG
HRELQIFSERTRHQ	SKEWVYECEPQRAA	KKASEPRDNFKALS
KCQVPTVWIAAFHQ	VIYDMEYSECYSVG	VIVPRINEHYHKFN
NNLMLKVTSNLSRD	WSYSEYVDTLENAE	VTVSALGTDTVASV
PNTYERDAFNST	DQVLNPISKAKLDK	SNYEGRDVHPIANA
KQAPKEKGYSKDEW	RDQARQSVPCIRVK	VHADTGGNVRGVKY
SERFKTSKLLTSEY	YDDDWDNRKKAYTKP	HRLDWNEGMYSANA
EKAIAQPRYSHRPM	NVDQSVKYHRVHSD	NTVSSNVGKQFNYI
LDEEVFTFENLRIY	SMSTQAAGNECAEI	WDPFCGEWKHAPDY
KDGREKQDPEFTIT	TKPKEDKDKVKLSA	RSPQDMLTGFWLAH
NEPLESAVDVYTKS	RTSCSEDIDATRKN	THEDFRYDDVFEGN
FCTVLKMRFPQQWR	SHWNKRSLGAQPYY	FVPWDTKRIFTTVD
IPQWWQQDHETAPF	NTDTSVKGNMAKD	RFRGTCPMWKCHFL
YWVWKILIYYAPHL	PVHYDFTFELILRM	EHSDWVIFWGIHNT
AELKLNKDVRGYFF	IMGWRVTRNYDVHH	DSTSGPVKTKAPVQ
RSRRKAEFMHAFTQ	VRKVRTWRCPVVDG	KHTGMAFKPVEYFW
LDTFFSKVWYNLYH	TPNFIPLPFYFYGR	WKRHYWQTTPMFMT
NFSDAVYIVRIIIS	SPDMKNTNVAYIGS	TDVLPNYIKAMNGH
PYFNQLQKDRSDLN	HMKHYEERERVVWI	DQERCEQPDRRIKD
DAAHKHHRDLGGQH	KSTHHESWLKIVDY	DLICFSTVVRPFFG
SEVSWKENVPDEDD	LAEVDHYMRGQDPR	RSSYHGKYQSPMS
NDQKREKCRGLPLG	NYLYNEILSMNKFR	PERNSTLPPNYLIS
SGFPDKIEFPTQDC	TQELYTDPYERITH	SHHFNWMNLMAVQY
LPTATGNMQSVPLI	YLPVFAHYVDSCDA	ESTFVYQENDLRN
RELLKCLDQMGVNY	TAATKLHRPIRRGS	ARDLISIWWMPQGP
QQRKTRVKEKRALG	QKDTSRKNNKEQGL	SVLYKPMGNARQR
QAAYFHGERSAAGS	KEEKSEEEHEGFRD	RQNRGIERRAAKLG
PSPKDSSLRPATSF	KFWDGHRHAMKEMS	QTEDEVKRNPNIIEV
RSIEGDSYPKSNDN	NRGLTDAYGAETRL	PAYRDGRDKLCSPK
KMIVQDIAWPWYGS	HSDKTIKTYTKMLR	RIQRHSQTTVEEDD
STTIDQGSQMKSPV	KFREAGLFHNGIAR	DPFKRGSGPKNDLL
DDEDSTSENNLEKK	TQPRKTAMINWFYK	FRALKYTQWERHDE
SFSWYRDFS RPNIN	DCQLWYLQQKIDKF	NVFNSYGPLTTFSH
DDKLGKGSPIKSFD	YPKNCIVLLFEWQD	FQETGRFHTKSVCA
GETRNPMMWLVGAG	PDNMKTRIVAGIAR	KNRADQTTSPHFWA
EVWTPPLDHFPPWHY	HLLEGHFVREEQH	VTPRSNTPLRVQLK
RHFFQVWNIRQCEN	ESQIDQGGWLSEHI	GVRHDLWNDAPS NF
NTFLHANCWVAIHG	STDDYEDAASQGTE	IDRKWPKTEHTRPD
KRMRTINFRSWNAQ	YFMVIYFALISAYD	NVQGD FWNVS YIS
DGSSLSQTYNPEYY	DVCSWTSSDMTNMG	RQLGQPPLLSTFPE

YRRKKRKQKYNQKV	QMPGLMYHTMAHNF	KPSASNYWKDWIER
NTMNVHLGCDFPFL	EMKDAFGHLHGIRC	LFKTDGRGVAHRTMD
FYHMF IQQHQLPW	RRQIKQTLDSDEVY	QHDNTYRHLAWDYF
LVTDARTKYKEEGV	IYKLYRSACGALRD	AHPKWGNVLSFFNA
ANHYKDGESEKAN	TRRYNIMSKIFNTW	VGSTGSDYNQTHY
MYACNKALKNFTPN	HSRDWATIRYCTNN	REDLRIKRPEEEGQ
EIWASYFVDIRSMY	CDVKDIGFGLLQPG	KHDKPENDVVYQKL
FRTETKRAQQHPTG	NSFEVGDGPNEIYD	YGEDHRRDKKPGQE
TDSQLWYLQQKIDK	KYEDDGVNRRTKND	LFSNGTVTRPKYSY
RTDTAFAPVNPAGWT	KTNYTNTAKNVKSI	RPEIHAGSARFSED
FYQQQFWQRFSECA	YSDKKTYTDTRPRQ	MDGRPESPSTMAKW
KAWSDMRLLPHEFK	SVQFWFRQGREQTG	RKSYNKRDKTGPE
VRQVQRSKKMHKKG	KQNQSRVPNQISLG	GGDGGGEQDMKVQY
MNDPSQVSSFLKKT	QHPEGDQNRSTRGP	KGTINFVADHHPDN
NKAGWVVELAAKVY	GYLVDDHAGLIATY	KIIQAKFHPNWDNH
VRFTLMEITTYTVMS	AFYLRTAHKPVVHR	EPMRHSIRGWNQSQ
EFFLMFEDWKWENL	PNRLQDGKQWNTSN	EYRKLTQWYQPEE
DTDAGTLTAQGSRH	INLATSDTEDSPGN	LLVEITRLKGHKWY
QEYLGGLNIAITFQH	VEQVVSRRKFGFWKF	QSFTQKNNAIQFNK
VVTIKTITTKDMVN	LNMPKSENYCRIRV	KAHLNLNLRNFNCT
TGGPWLAGREWAPG	ALFIFAERDPMKGV	RLSATENELRQGRP
HDDEFQPTWVPNTE	LTKFENNETDSSNA	AKNENTRYFGNYTG
SRRPMHHPDDPHT	VPFEFEIDFAQAQN	VDEEREGPQLERHC
TQISFVVPYNSPLS	IHAKYVAKFINNLV	YEKNGTRKRPRGAV
NWTCKNQVIWEEMS	RYPANDWRTNSEEI	SRAGPFAAPKVPAD
VDFVAEPVKLPENQ	VLRHEHSYTESSRY	PPNEWMDDWQPGEN

Die 255 ausgewerteten Peptidsequenzen der Glaträger.

# Abkürzungen

Abkürzung	Bedeutung
Ala	Alanin
Ak	Antikörper
Arg	Arginin
AS	Aminosäure(n)
Asn	Asparagin
Asp	Asparaginsäure
bzw.	beziehungsweise
Cys	Cystein
CV	Variationskoeffizienten (von engl. <i>coefficient of variation</i> )
d.h.	das heisst
DMSO	Dimethylsulfoxid
DRFZ	Deutsches Rheuma Forschungszentrum
dt.	deutsch
DZ	Diagnostische Zuverlässigkeit
dpi	days post infection (dt. <i>Tage nach Infektion</i> )
engl.	englisch
FDR	Rate der falsch entdeckten, von engl. <i>false discovery rate</i>
Gln	Glutamin
Glu	Glutaminsäure
Gly	Glycin
<i>H. polygyrus</i>	<i>Heligmosomoides polygyrus</i> (Fadenwurm)
His	Histidin
IgA	Immunglobulin A
IgE	Immunglobulin E
IgG	Immunglobulin G
IgM	Immunglobulin M
Ile	Isoleucin
Leu	Leucin
Lys	Lysin

Met	Methionin
<i>MHC</i>	Haupthistokompatibilitätskomplex (engl. <i>Major Histocompatibility Complex</i> )
N	Anzahl
oPfp	ortho-Pentafluorphenyl
PAM	'Prediction Analysis for Microarrays'-Software
Phe	Phenylalanin
Pro	Prolin
<i>RAD</i>	<i>Receptor Affinity Distribution</i>
Rep.	Repetitionen
Res.	Resultate
ROC	von engl. <i>Receiver Operating Characteristic</i> , ROC-Kurve ist Darstellung der Abhängigkeit der Effizienz von Fehlerrate
SAM	'Significance Analysis of Microarrays'-Software
sas	solvent accessible surface (dt. <i>flüssigkeitszugängliche Oberfläche</i> )
Ser	Serin
SI	Signalintensität, Signalintensitäten
SNR	Signal-Rausch-Verhältnis (von engl. <i>signal-to-noise ratio</i> )
TBC	Tuberkulose
TFA	Trifluoro Acetic Acid (dt. <i>Trifluoressigsäure</i> ), F <sub>3</sub> CCOOH
Thr	Threonin
T-PBS	Phosphate buffered saline (dt. <i>Phosphat-gepufferte Salzlösung</i> ) mit 0.05 % Tween 20
Tris	Trishydroxymethylaminomethan
Trp	Tryptophan
T-TBS	Tris buffered saline (TRIS:=Tris(hydroxymethyl)-aminomethan gepufferte Salzlösung) mit 0.05 % Tween 20
Tyr	Tyrosin
Val	Valin
z. B.	zum Beispiel



# Danksagung

Für Unterstützung und Hilfe danke ich:

Meiner Betreuerin **Michal Or-Guil** für ihre Fachkenntnis, Unterstützung und wissenschaftliche Begeisterung. Danke Dir auch für die Freiräume und Möglichkeiten, dank derer ich trotz komplizierter Umstände diese Arbeit beenden konnte. Danke auch für die Ermunterung und die finanzielle Unterstützung zum Besuch der vielen Konferenzen und Kurse.

**Hanspeter Herzel** und **Sepp Hochreiter**, für Euer Einverständnis, diese Arbeit zu begutachten.

Meinen Kollegen aus der Arbeitsgruppe '**SysImm**', allen voran **Juliane Bongartz** für die gute Zusammenarbeit, für Inkubationen, Datenmanagement, Korrekturlesen, und vieles mehr; **Armin Weiser** für IT-Dienste und mathematischen Support; **Nicole Wittenbrink** für äusserst hilfreiches Korrekturlesen nicht nur dieser Arbeit und für die Bereitschaft, uns jederzeit mit ihrem präzisen, immunologischen Wissen und Denken zu unterstützen; der ganzen Gruppe danke ich für Diskussionen und Unterstützung und alles, was eine Arbeitsgruppe ausmacht, was ich im letzten Jahr für lange Zeit sehr vermisst hatte.

**Mike Schutkowski** für viele interessante Diskussionen und Literaturtipps; **Ulrich Reineke** für gute Diskussionen, Informationen und Ratschläge, vor allem zu Anfang der Arbeit, sowie auch **Robert Carus** und **Jonny Zerweck** für viel Unterstützung und Mitdenken bei dieser Arbeit (alle von **JPT Peptide Technologies GmbH**, Berlin).

Der Gruppe '**Molekulare Bibliotheken**', vor allem **Bernhard Ay**, für die Synthese der Membranen auf dem Autospotti, für gute Diskussionen und guten Informationsaustausch, **Astrid Leichsenring** und **Christiane Landgraf** für die Synthese der manuell hergestellten Membranen, **Victor Tapia**

für Inkubationen von Membranen und Glaträgern, die teilweise auch in dieser Arbeit ausgewertet wurden, und **Rudolf Volkmer** für das Ermöglichen der Zusammenarbeit und die Erlaubnis der Nutzung der Labors, Geräte und Reagenzien.

Der Arbeitsgruppe von Susanne Hartmann aus der '**Molekularen Parasitologie**', HU Berlin, vor allem Michal Sereda und Sebastian Rausch für die Seren von gesunden und *H.-polygyrus*-infizierten Mäusen.

**Jorge Carneiro** für sehr gute Diskussionen und seine ansteckende Begeisterungsfähigkeit.

**Johannes Schuchhardt** für gute Diskussionen und Unterstützung.

**Claudia Berek** vom DRFZ für gute Diskussionen über Antikörperrepertoires und deren Entstehung sowie für Unterstützung und Beratung bei der Organisation von Seren.

**Peter Bühlmann** und dem **Seminar für Statistik** der ETHZ danke ich dafür, dass sie mir in den letzten Monaten derart unbürokratisch und herzlich einen Arbeitsplatz zur Verfügung gestellt haben und mir damit ein Umfeld geboten haben, um grosse Teile dieser Arbeit zusammenzuschreiben.

**Mike Niederer** für grossartige Hilfe beim Korrekturlesen und beenden dieser Arbeit.

**Andrea Trueb** fürs last-minute Korrekturlesen.

**Jutta Rodemann** fürs Korrekturlesen und für ein »zweites Zuhause« in Berlin.

Meiner Mutter **Brigitte** dafür, dass sie in den letzten 16 Monaten jede Woche einen Tag mit Leonard verbracht hat, eine sehr geschätzte Erleichterung, das Vorliegende zu schreiben. **Marco** danke ich für sehr ausführliches Korrekturlesen. Meinen Schwestern: **Francine** und **Julia** für öfteres Babysitten, und **Lea** für die gute Absicht.

Meinen Männern **Ben** und **Leonard** für ihre Geduld und Ben zusätzlich für seine Unterstützung.

Diese Arbeit wurde freundlicherweise von der VolkswagenStiftung  
unterstützt.





# Selbständigkeitserklärung

Hiermit erkläre ich, dass die vorliegende Arbeit mit dem Titel 'Informationsgewinnung mittels Bindungsanalysen von Serumantikörpern an Peptidbibliotheken' selbständig von mir angefertigt wurde. Sämtliche Hilfsmittel, Hilfen sowie Literaturquellen sind als solche kenntlich gemacht.

Ich habe mich anderweitig nicht um einen Doktorgrad beworben und besitze auch keinen entsprechenden Doktorgrad.

Mir ist die dem Verfahren zugrunde liegende Promotionsordnung bekannt.

Zürich, den

Nicole Bruni